AWARD NUMBER: W81XWH-10-1-0404

TITLE: Receptive Vocabulary Knowledge in Low-Functioning Autism as Assessed by Eye Movements, Pupillary Dilation, and Event-Related Potentials

PRINCIPAL INVESTIGATOR: Barry Gordon, M.D., Ph.D.

CONTRACTING ORGANIZATION: Johns Hopkins University
Baltimore, MD 21218

REPORT DATE: JUNE 2014

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
June 2014

2. REPORT TYPE
Final

3. DATES COVERED
1Jun2010 - 30Jun2014

4. TITLE AND SUBTITLE

Receptive Vocabulary Knowledge in Low-Functioning Autism as Assessed by Eye Movements, Pupillary Dilation, and Event-Related Potentials

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-0404

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Barry Gordon, M.D., Ph.D.
Kerry Ledoux, Ph.D.

E-Mail: bgordon@jhmi.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Johns Hopkins University
3400 N Charles St W400 Wyman Park
Building
Baltimore, MD 21218

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We have been testing the hypothesis that relatively implicit measures of cognitive processing (eye movements, pupillary dilation monitoring, and the N400 component of event-related potentials) will prove sensitive to receptive vocabulary knowledge, even in the absence of more traditional behavioral responses. We have sought to first demonstrate the use of these measures in three populations in whom behavioral responses are expected to be reliable: normal adults, normally developing children, and higher-functioning individuals with autism. In all three groups, the implicit measures differentiated known from unknown words: eye movements were faster to a named picture for known words; pupillary dilation from baseline was greater in the unknown condition; and an N400 congruency effect was observed for known (but not unknown) words. Our results also suggest that these measures similarly differentiate known from unknown words in lower-functioning individuals with autism, even in the absence of a behavioral response. These results suggest that these measures may be used as valid measures of comprehension, even in nonverbal, non-responding individuals.

15. SUBJECT TERMS
autism, lower-functioning individuals, vocabulary, event-related potentials, eye movements, pupillary dilation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	141	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

1. INTRODUCTION

Approximately 50% of individuals affected by autism fail to develop useful speech, and many of these individuals never learn to communicate in any functional way. An important scientific and practical question about such individuals, as well as in those with other diagnoses and a similar inability to express themselves, is whether this lack of expressive ability is necessarily accompanied by an equally severe deficit in knowledge of receptive language. Little rigorous research has been directed at this possibility, both because of the difficulty of working with such low-functioning subjects and because of the lack of sensitivity of most traditional behavioral methodologies. Recently, however, several experimental methodologies have been developed and refined to the point that they may prove sensitive enough to provide reliable evidence of comprehension, even in the absence of more traditional behavioral responses such as speech and gesturing, and even at the individual subject level. We have been developing the use of three such research methods to attempt to detect receptive vocabulary knowledge – eye movement recording, pupillary dilation monitoring, and event-related brain potentials. We have been testing whether these relatively implicit measures of comprehension actually do reflect single-word comprehension in participants in whom we expect reliable behavioral responses to serve as comparison measures (normal adults, normally developing children, and high-functioning individuals with autism), as well as in low-functioning, nonverbal individuals with autism, for whom overt behavioral responses might be unreliable or even impossible.

2. KEYWORDS:
autism, lower-functioning individuals, vocabulary, event-related potentials, eye movements, pupillary dilation

3. ACCOMPLISHMENTS

What were the major goals of the project?

Experiment 1: Validating the use of the eye movement (EM), pupillary dilation (PD), and event-related potential (ERP) techniques for the measurement of receptive vocabulary knowledge – normal adults (Planned: Months 1 – 6; actual: Months 1 - 12)

- 1a. Data collection
- 1b. Data analysis
- 1c. Manuscript preparation

Experiment 2: Validating the use of the EM, PD, and ERP techniques for the measurement of receptive vocabulary knowledge – normally-developing children (Planned: Months 1 – 12; actual: Months 6 - 24)

- 2a. Participant recruitment – to be continued and elaborated from initial efforts
- 2b. Data collection
- 2c. Data analysis
- 2d. Manuscript preparation

Experiment 3: Validating the use of the EM, PD, and ERP techniques for the measurement of receptive vocabulary knowledge – high-functioning individuals with autism (Planned: Months 1 – 18; actual: Months 6 - 48)

- 3a. Participant recruitment – to be continued and elaborated from initial efforts
- 3b. Autism diagnosis verification via administration of the ADOS and ADI-R
- 3c. Data collection
- 3d. Data analysis
- 3e. Manuscript preparation

Experiment 4: Extending the use of the EM, PD, and ERP techniques for the assessment of receptive knowledge to low-functioning individuals with autism (Planned: Months 1 – 24; actual: Months 6 - 36)

- 4a. Participant recruitment – to be continued and elaborated from initial efforts
- 4b. Autism diagnosis verification via administration of the ADOS and ADI-R
- 4c. Acclimation to eye-tracking and ERP equipment
- 4d. Individualized selection of stimuli
- 4e. Data collection
- 4f. Data analysis
- 4g. Manuscript preparation

Experiment 5: Using the EM, PD, and ERP techniques to study new word learning in low-functioning individuals with autism (Planned: Months 1 – 36; actual: Months 45 - 48)

- 5a. Individualized selection of stimuli for exposure and non-exposure sets
- 5b. Learning period
- 5c. Post-test data collection
- 5d. Data analysis
- 5e. Manuscript preparation

What was accomplished under these goals?

Experiment 1: Validating the use of the eye movement monitoring (EM), pupillary dilation monitoring (PD), and event-related potential (ERP) techniques for the measurement of receptive vocabulary knowledge – normal adults

Our first experiment was designed to validate the use of the three implicit methodologies to detect receptive vocabulary knowledge in normal adults, a participant population in whom overt behavioral responses would be expected to be reliable (and thus capable of serving as a measure of comparison). Participants were asked to engage in two separate tasks using the same set of 160 words and pictures. Eighty of the words were very high frequency and were expected to be very familiar to all of the adults; these included words such as *airplane* and *camera*. The remaining 80 words were low frequency, relatively unfamiliar words that were not expected to be known by many of the participants (as confirmed by prior pre-testing). Examples of words in this set included *agouti* and *cainito*. All words were concrete and highly imageable. High-resolution, color digital pictures were selected to represent each word. In the *forced-choice recognition task*, participants were asked to use the mouse to select one of four pictures presented simultaneously on a computer screen after hearing one of the objects named. We simultaneously collected eye movement and pupillary dilation data using an ASL Model 504 eye-tracking system. In the *congruity task*, a picture was presented on the computer screen, accompanied by the auditory presentation of a single word, which either matched (congruous condition) or did not match (incongruous condition) the pictured item. Participants were asked to push a button to indicate whether the auditory word and the picture matched. Simultaneously, ERPs were recorded using Electrical Geodesics Inc.'s 256-channel Hydrocel Geodesic Sensor Nets. Finally, normal adult participants were asked to participate in a word familiarity post-test, in which they were asked to rate their familiarity with the 160 words used in the experiment, on a scale from 1 (very unfamiliar) to 9 (extremely familiar), with an additional option of 0 (no familiarity whatsoever).

During the third year, we improved the sophistication of our data analysis techniques even further. With the hiring of a new postdoctoral associate, we were able to develop methods to apply principal components analysis (PCA) and independent components analysis (ICA) to our EEG data. These analysis techniques allow for the identification of eye blink, eye movement, and other artifacts in the data, and importantly allow the removal of these artifacts from the EEG without the loss of the entire data set for that particular trial (as had often been the result under our previous techniques). That is, once the activity associated with these artifacts (and with other non-cognitive activity) is identified, it is mathematically isolated and removed, with the electrical activity specific to the cognitive event of interest remaining intact. These techniques thus allow the retention of a much larger proportion of the data from most participants, and this is especially true for those participants who have difficulty performing the experiment without creating a lot of artifacts, such as children or individuals with autism. We believe, then, that these techniques will greatly improve our ability to analyze larger numbers of clean, artifact-free trials from all participant groups.

This has proven true in the normal adult data analysis. Re-analyzing data using these techniques, our final analysis was able to include data from seven additional participants whose data had previously been ruled unusable because of the loss of a large number of trials due to

artifacts. Subsequent manuscript drafts included this larger group of participants. For this group, the main findings that we had previously reported held true: eye movements to the picture that matched the auditory word were faster for known than for unknown words. End-of-trial fixations were also on the named picture more frequently for known words. Pupillary dilation from baseline was greater in the unknown condition (evidencing the greater engagement of cognitive resources when the word is unknown). Additionally, an N400 congruency effect was observed in the event-related potentials for known words, but not for unknown words. Thus, all three implicit measures (EM, PD, and ERPs) were able to distinguish the processing of known from unknown words in this participant population. The manuscript with accompanying data describing these results has been published in *Behavior Research Methods* (see Ledoux, et al., 2015 in Appendix 1).

During our third year, we also developed a model that allows us to predict the knowledge ratings provided by the normal adults based on the results from the implicit measures. We used the EM, PD, and ERP results jointly to create a regression model that predicts participants' explicit word knowledge ratings. The predicted knowledge ratings from the model were then used to recode words as "known" or "unknown." In this way, we used the implicit measures to provide us with information about which words are truly likely to be known or unknown to a given individual participant, and re-coded all of the stimulus items specifically based on that information, for each participant. We then looked at the ERP effects under the individualized coding scheme. Stronger differences were observed on the N400 component; specifically, the N400 to the congruent picture-word pairs that were known to the participant showed a larger reduction in amplitude relative to those that were known but incongruent. The amplitude of the N400 to words that were unknown was intermediate to the two known conditions, and did not differ by congruency (as would be expected for words about which participants truly have no knowledge). In this way, the regression model allows us to use the data to determine which words are most likely to be known or unknown to each participant individuallyin a way that does not rely on overt behavioral responses. We are revising a manuscript describing this modeling work after receiving reviews back from *Behavioral Research Methods* (see Coderre, et al., under revision, in Appendix 3). We hope that this model will prove useful in further analyses of the data from the typically developing children and the participants with autism, in whom there is expected to be greater variability in knowledge about the vocabulary words and in whom behavioral responses are not always the best indicator of that knowledge.

Experiment 2: Validating the use of the EM, PD, and ERP techniques for the measurement of receptive vocabulary knowledge – normally developing children

Our second experiment was designed to validate the use of the three implicit methodologies for detecting receptive vocabulary knowledge in normally developing children (ages 5 – 17), another participant population in whom overt behavioral responses would be expected to be reliable (and thus capable of serving as a measure of comparison). The child participants were tested on the Peabody Picture Vocabulary Test (PPVT; [1]), the Kaufman Brief Intelligence Test (KBIT; [4]) and the Autism Spectrum Screening Questionnaire (ASSQ; [2]), the latter of which was used to ensure that none of the normally developing children exhibited excessive behaviors associated with autism. All of the children were asked to complete the forced-choice recognition task and the congruity task described above. Older children (those old

enough to understand and properly perform the task; generally ages 10 and above) were also asked to complete the familiarity post-test described in Experiment 1.

Our preliminary analyses on the data from 20 children demonstrate that the results for the child participants are very similar to those observed for the normal adults. Behaviorally, children were faster and more accurate at both the forced-choice task and the congruity task for known words than for unknown words. Eye movements to the picture that matched the auditory word were faster for known than for unknown words. End-of-trial fixations were also on the named picture more frequently for known words. Pupillary dilation from baseline was greater in the unknown condition. Additionally, an N400 congruency effect was observed in the event-related potentials for known words, but not for unknown words. Thus, all three implicit measures (EM, PD, and ERPs) were able to distinguish the processing of known from unknown words for the normally developing children that were tested (see Gangopadhyay, et al., 2012, in Appendix 4).

We have continued to re-analyze the children's data using the PCA/ICA techniques described in Experiment 1 with the hope that we can retain more trials from the participants and thus strengthen our results. We have also begun to apply the regression model described in Experiment 1 to the data of the typically developing children to reclassify individual stimuli as "known" or "unknown" for each individual child, based on the results from the three implicit measures.

Experiment 3: Validating the use of the EM, PD, and ERP techniques for the measurement of receptive vocabulary knowledge – high-functioning individuals with autism

Our third experiment was designed to validate the use of the three implicit methodologies to detect receptive vocabulary knowledge in high-functioning individuals with autism, another participant population in whom overt behavioral responses would be expected to be reliable (and thus capable of serving as a measure of comparison), but which also offers a more closely-matched comparison group to the low-functioning individuals with autism. Participants were administered the Kaufman Brief Intelligence Test (KBIT; [4]), the Autism Diagnostic Observation Schedule (ADOS; [5]), and the Autism Diagnostic Interview – Revised (ADI-R; [6]) to confirm diagnosis and to determine level of functioning/verbal ability.

We recruited higher-functioning individuals with autism throughout the time of the grant, and throughout the time of the no-cost extension, as these participants proved to be the most difficult for us to recruit. We suspect that for this group in particular, participation in experiments might be less appealing. There are many reasons why this may be true. For instance, higher-functioning individuals may be intrinsically less motivated to help in the process of discovering potential deficits in autism and in developing intervention strategies because they see less need for these strategies for themselves, as they are often able to function quite well in school or work settings. Alternatively, higher-functioning individuals may be too busy to participate in research because they attend more mainstream schools or are involved in afterschool activities, making it difficult for them to find the necessary time for testing. We have tried to identify such reasons and have made attempts to circumvent them in our recruitment efforts. Ultimately, we were able to collect complete data from approximately 15 high-functioning (primarily adult) participants.

Among this group, our results show similarities to those observed for normal adults and for typically developing children. The individuals in this group were able to make reliable

behavioral responses. Behaviorally, they were faster and more accurate at both the forced-choice task and the congruity task for known words than for unknown words. Eye movements to the picture that matched the auditory word were faster for known than for unknown words. End-of-trial fixations were also on the named picture more frequently for known words. Pupillary dilation from baseline was greater in the unknown condition. Additionally, an N400 congruency effect was observed in the event-related potentials for known words, but not for unknown words. Thus, all three implicit measures (EM, PD, and ERPs) were able to distinguish the processing of known from unknown words for the high-functioning individuals with autism that we have tested. Corresponding data has been included in recent poster presentations. (see Coderre, et al., 2014, in Appendix 5 and Coderre, et al., 2015, in Appendix 6). We have also begun to apply the regression model described in Experiment 1 to the data of the high-functioning individuals with autism, to reclassify individual stimuli as "known" or "unknown" for each participant, , based on the results from the three implicit measures.

Experiment 4: Extending the use of the EM, PD, and ERP techniques for the assessment of receptive vocabulary knowledge to low-functioning individuals with autism

Our fourth experiment was designed to extend the use of the three implicit methodologies to detect receptive vocabulary knowledge to a population in whom behavioral responses are generally less reliable (or absent altogether) – low-functioning, low-verbal or nonverbal individuals with autism. Participants were administered the Kaufman Brief Intelligence Test (KBIT; [4]), the Autism Diagnostic Observation Schedule (ADOS; [5]), and the Autism Diagnostic Interview – Revised (ADI-R; [6]) to confirm diagnosis and to determine level of functioning/verbal ability.

We tested approximately 25 low-functioning individuals with autism, from whom we received complete and usable data from 10. Reasons for data exclusion were similar to those described for the other groups. Additionally, even with acclimation training, low-functioning participants have a much harder time engaging in the tasks for extended periods of time. Therefore, all of the eye-tracking and ERP artifacts for this group are very pronounced. The behavior of individuals in this population is quite variable, so that on some days, they are unwilling to participate at all. Also, participant attrition is a problem, given the large time commitment required of the participants and their families for successful acclimation training and testing.

All participants were minimally verbal to nonverbal. Stimuli for the low-functioning group were drawn from the larger pool of 160 words and pictures, but were individualized for each participant based on parental/caregiver reports of items that were expected to be known receptively by the participant. Parents/caregivers were asked to complete the MacArthur-Bates Communicative Development Inventory – Words and Gestures ([3]), plus a similar experiment-specific inventory that covered those words from our set of 160 that were not included on the MacArthur-Bates. These measures thus provided information about what words were likely to be known (and unknown) receptively by the individual. The number of stimuli tested were determined for each individual to maximize signal-to-noise ratio while minimizing experiment length. For some individuals, for whom the pool of known words was small, repetition of items within a testing session, or the repeated testing across multiple testing sessions, was necessary to adequately assess their receptive knowledge.

In addition to the assessments provided by the ADOS and the ADI-R, each low-functioning participant received a series of behavioral assessments designed to evaluate his/her ability to successfully participate in our language testing. We assessed potential participants on things such as their ability to sit still for extended periods of time; their ability to look at the computer screen; their ability to tolerate the eye tracking and ERP equipment; and their likelihood to exhibit adverse behaviors (such as hitting, biting, or other aggressive behaviors). Based on these assessments, an individual determination was made as to the appropriateness of participation and the need for further individualized training to acclimate the participant to the eye-tracking and ERP equipment and experiment procedures. Such training was then conducted as needed over a period of days or weeks in our testing space and at the participant's home.

After training, participants completed the same forced-choice and congruity tasks as described for Experiments 1-3. However, they were not required to make any overt behavioral response using the mouse or pressing a button. (Some low-functioning individuals with autism are very familiar with computer programs of the type used in our experiments and would like to engage in some kind of task during the experiment. These participants are allowed to make responses as they wish. However, importantly, the successful analysis of the implicit measure data in this experiment does not depend upon the behavioral completion of these tasks.)

Our results from the low-functioning participants show a fair amount of individual variability, but the results for most trials across participants show great similarities to the results from our other participant groups: eye movements were faster and more accurate for known words than for unknown words. Changes in pupillary dilation were greater to unknown than to known words. Finally, several of the lower-functioning participants showed evidence of an N400 congruency effect, with a larger amplitude in the N400 time range in the incongruent condition relative to the congruent condition, but only for the words that were expected (based on parental report) to be known by the individuals. The data for this group are currently included in a manuscript that has been submitted for review to *Journal of Speech, Language, & Hearing Research*. (see Coderre, et al., under review, in Appendix 2).

We have also begun to apply the regression model described in Experiment 1 to the data of the lower-functioning individuals, to reclassify individual stimuli as "known" or "unknown" for eachparticipant individually, based on the results from the three implicit measures. This will be especially important in this group, as these individuals cannot provide very accurate information about what they know about words themselves. Relying on parental/caregiver report is not necessarily accurate either, as it is certainly possible (in fact, from our results, expected) that these individuals know more about words than they can demonstrate, and others may not have a complete sense of what these individuals do and do not know (see Coderre, et al., under revision, in Appendix 3).

Experiment 5: Using the EM, PD, and ERP techniques to study new word learning in low-functioning individuals with autism

Our fifth experiment was designed to examine changes in EM, PD, and ERP measures in nonverbal, low-functioning individuals with autism that accompany repetitive exposure to new words during a learning period.

During year two, we completed pilot testing to explore different teaching methods and stimulus sets for this phase of our experiment. In year three, we enrolled our first participant for actual testing in Experiment 5. The participant that we enrolled, DL, was a 22-year-old

functionally nonverbal male with autism who had successfully completed Experiment 4 and with whom we have worked extensively in the past five years. This individual was generally quite tolerant of the eye tracking and ERP equipment and seemed well suited to further and repeated testing in Experiment 5. Over a period of six months, we worked with him and his family to select appropriate stimuli for the training study and to optimize our proposed training procedure.

Unfortunately, at about the time that we were hoping to begin the actual word training with DL, his parents decided to temporarily withdraw him from the study. DL recently transitioned from a school setting to an adult activity center, and this change resulted in increases in obsessive-compulsive behaviors (which were previously present, but have since become worse) and other anxiety-related behaviors. His parents worried that the word training sessions and the post-training testing would add too much novelty to his already altered world. They requested that we postpone the testing for at least six months.

What opportunities for training and professional development has the project provided?

Nothing to report.

How were the results disseminated to communities of interest?

In addition to the journal article manuscripts that are currently in press, under review, or in preparation, the results of these studies were shared with the scientific community through several presentations at academic conferences:

Coderre, E., Chernenok, M., Bosley, L., O'Grady, J., Gordon, B., & Ledoux, K. (2015, May). Implicit Measures of Receptive Vocabulary Knowledge in Low-Functioning Individuals with Autism. Poster presented at the 14th Annual International Meeting for Autism Research, Salt Lake City, UT.

Coderre, E., Bosley, L., Chernenok, M., Gordon, B., & Ledoux, K. (2013, November). Modeling Implicit Measures of Receptive Vocabulary Knowledge in Normal Adults. Poster presented at the 54th Annual Meeting of the Psychonomic Society, Toronto, Canada.

Gangopadhyay, I., Ledoux, K., Bosley, L., & Gordon, B. (2012, May). Assessing Receptive Vocabulary Knowledge in Individuals with Autism Using Implicit Measures. Poster presented at the 11th Annual International Meeting for Autism Research, Toronto, Canada.

Gangopadhyay, I., Ledoux, K., Bosley, L., & Gordon, B. (2012, April). The Use of Implicit Measures to Assess Vocabulary Knowledge in Normal Adults and Normally Developing Children. Poster presented at the 19th Annual Meeting of the Cognitive Neuroscience Society, Chicago, IL.

Gangopadhyay, I., Ledoux, K., Bosley, L.V., & Gordon, B. (2011, November). The Use of Implicit Measures to Assess Receptive Vocabulary Knowledge in Individuals with Autism. Poster presented at the 3rd Annual Neurobiology of Language Conference, Annapolis, MD.

Ledoux, K., Pickett, E.J., Van Droof, L.V., Buz, E., Billings, N.M., & Gordon, B. (2010, November). Receptive Vocabulary Knowledge in Individuals with Autism as Assessed by Eye Movements, Pupillary Dilation, and Event-Related Potentials. Poster presented at A Brain Research Meeting: The Emerging Neuroscience of Autism Spectrum Disorders, San Diego, CA.

What do you plan to do during the next reporting period to accomplish the goals?

Nothing to report (final report).

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

One of the important ways in which this work has had an impact is through its inclusion of minimally verbal or nonverbal, low-functioning individuals with autism. Such individuals have been woefully under-represented in prior research. The inclusion of such individuals in empirical studies of language and cognitive processing has been difficult for practical reasons – the absence of a reliable verbal or behavioral response makes the accurate assessment of these individuals extremely challenging. Additionally, many of these individuals exhibit behavioral tendencies that often preclude their participation in research settings. However, an understanding of cognitive processing in low-functioning individuals with autism is critical to our understanding of this condition; the current over-representation in empirical studies of autism of higher-functioning participants does not allow a full understanding of the cognitive deficits (nor, for that matter, of the preserved cognitive abilities) in individuals with autism as a whole. Thus, our research has contributed important knowledge about the cognitive capabilities of individuals who are often excluded from research in autism.

Even more specifically, our studies have demonstrated that even those individuals with autism who are minimally verbal or nonverbal may still have intact verbal comprehension abilities. The use of implicit measures in our studies has shown evidence of comprehension markers in low-functioning participants that are very similar to those observed in verbal populations (normal adults, typically developing children, and high-functioning individuals with autism). There have long been strong suspicions, by those in close contact with nonverbal individuals with autism (such as parents and teachers), that these individuals often “know” more than they can express. Our data, which used parental report as a standard for what was known to the participants , suggest those suspicions may have merit.

The demonstration that individuals with autism who are unable to make reliable behavioral responses nonetheless exhibit implicit evidence of word knowledge has appreciable scientific implications and even more significant practical implications. Scientifically, it is a well-documented principle in the acquisition of language in normally-developing children that comprehension precedes expression: young children almost always show evidence of being able to understand a word’s meaning before they can produce the word. To the extent that this principle also applies in autism (which, to this point, is an open question), the demonstration of receptive abilities in nonverbal individuals could lay an important foundation for better understanding their baseline communication and comprehension abilities.

What was the impact on other disciplines?

As a practical matter, knowing that an individual can understand language even when he or she does not speak can support the development of more intensive speech and language therapies, using a broader range of modalities, to capitalize on that individual's functional preferences or strengths. The ability to to document new word learning, without having to rely upon unreliable and insensitive behavioral measures, may be particularly important for this low-functioning individuals with autism. Teaching low-functioning individuals with autism is very difficult, in part, because their learning is typically slow, erratic, and difficult to detect by standard behavioral means. Our results may prove important in educational, therapeutic, and other clinical settings in which individuals with autism are taught new words.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

As described above, our results might have direct application in the fields of education, therapy, and learning with individuals with autism. To the extent that the recognition of comprehension abilities in minimally verbal to nonverbal individuals with autism can help inform better teaching or rehabilitation methods, these results could ultimately have direct benefits in helping to make the acquisition of language and communication easier for those individuals with autism who currently struggle most with such skills.

5. CHANGES/PROBLEMS

Nothing to report.

6. PRODUCTS

Publications, conference papers, and presentations

Ledoux, K., Coderre, E., Bosley, L., Buz, E., Gangopadhyay, I., & Gordon, B. (2015). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavior Research Methods, 48*, 285-305. (See Appendix 1.)

Coderre, E., Chernenok, M., O'Grady, J., Bosley, L., Gordon, B., & Ledoux, K. (Under review.) Implicit measures of receptive vocabulary knowledge in low-functioning individuals with autism. *Journal of Speech, Language, & Hearing Research*. (See Appendix 2.)

Coderre, E., Gordon, B., & Ledoux, K. (Under revision). The use of mixed-effects models to predict receptive vocabulary knowledge from implicit measures of language comprehension. *Behavior Research Methods*. (See Appendix 3.)

Coderre, E., Cherenok, M., O'Grady, J., Bosley, L, Gordon, B., & Ledoux, K. (2015, September). Event-Related Potentials as Implicit Measures of Vocabulary in Individuals with Autism. Poster presented at the American Neurological Association's 2015 Annual Meeting, Chicago, IL. (See Appendix 6.)

Coderre, E., Gordon, B., & Ledoux, K. (2014, April). Neural Connectivity During Semantic Processing of Pictures and Spoken Words in Autism Spectrum Disorders. Poster presented at the 21st Annual Meeting of the Cognitive Neuroscience Society, Boston, Massachusetts. (See Appendix 5.)

Gangopadhyay, I., Ledoux, K., Bosley, L., & Gordon, B. (2012, May). Assessing Receptive Vocabulary Knowledge in Individuals with Autism Using Implicit Measures. Poster presented at the 11th Annual International Meeting for Autism Research, Toronto, Canada. (See Appendix 4.)

Gangopadhyay, I., Ledoux, K., Bosley, L., & Gordon, B. (2012, April). The Use of Implicit Measures to Assess Vocabulary Knowledge in Normal Adults and Normally Developing Children. Poster presented at the 19th Annual Meeting of the Cognitive Neuroscience Society, Chicago, IL.

Gangopadhyay, I., Ledoux, K., Bosley, L.V., & Gordon, B. (2011, November). The Use of Implicit Measures to Assess Receptive Vocabulary Knowledge in Individuals with Autism. Poster presented at the 3rd Annual Neurobiology of Language Conference, Annapolis, MD.

Ledoux, K., Pickett, E.J., Van Droof, L.V., Buz, E., Billings, N.M., & Gordon, B. (2010, November). Receptive Vocabulary Knowledge in Individuals with Autism as Assessed by Eye Movements, Pupillary Dilation, and Event-Related Potentials. Poster presented at A Brain Research Meeting: The Emerging Neuroscience of Autism Spectrum Disorders, San Diego, CA.

Books or other non-periodical, one-time publications.
Other publications, conference papers, and presentations.
Website(s) or other Internet site(s)
Technologies or techniques.
Inventions, patent applications, and/or licenses.
Other products.

Nothing to report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?
Primary investigators:

Barry Gordon, MD, PhD
Dr. Gordon has overseen and contributed to the management of all aspects of the research reported herein. He has also contributed to the drafting of manuscripts and to the creation of conference presentations. He has worked approximately 12 person months over the four years of the grant.

Kerry Ledoux, PhD.
Dr. Ledoux has also overseen and contributed to the management of all aspects of the research reported herein, including participant recruitment and testing, stimulus generation and design, data analysis, as well as manuscript and conference presentation generation. She has worked approximately 24 person months over the four years of the grant.

Other personnel:
Post-doctoral fellow:

Emily Coderre, PhD
Dr. Coderre has contributed to the management of all aspects of the research reported within, including participant recruitment and testing, stimulus generation and experiment programming, data analysis, and drafting of manuscripts. She worked approximately 12 person months over the last two years of the grant (the time for which she was in our lab).

Research assistants:

Mariya Chernenok (approximately 27 person months over three years)
Ishanti Gangopadhyay (approximately 18 person months over two years)
Esteban Buz (approximately 9 person months over one year)
Nia Billings (approximately 9 person months over one year)
Laura Bosley (approximately 5 person months over four years)
Erin Pickett (approximately 4 person months over one year)

The research assistants, under the direction of the PIs and the postdoctoral fellow, assisted with all aspects of the research reported within, especially participant recruitment and testing, stimulus generation and design, and aspects of data coding and analysis.

Has there been a change in the active other support of the PIs?

Nothing to report.

What other organizations were involved as partners?

Nothing to report.

Appendix 1

Ledoux, K., Coderre, E., Bosley, L., Buz, E., Gangopadhyay, I., & Gordon, B. (2015). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavioral Research Methods, 48*, 285-305.

Behav Res
DOI 10.3758/s13428-015-0571-6

The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults

Kerry Ledoux · Emily Coderre · Laura Bosley · Esteban Buz · Ishanti Gangopadhyay · Barry Gordon

Abstract Recent years have seen the advent and proliferation of the use of implicit techniques to study learning and cognition. One such application is the use of event-related potentials (ERPs) to assess receptive vocabulary knowledge. Other implicit assessment techniques that may be well-suited to other testing situations or to use with varied participant groups have not been used as widely to study receptive vocabulary knowledge. We sought to develop additional implicit techniques to study receptive vocabulary knowledge that could augment the knowledge gained from the use of the ERP technique. Specifically, we used a simple forced-choice paradigm to assess receptive vocabulary knowledge in normal adult participants using eye movement monitoring (EM) and pupillometry. In the same group of participants, we also used an N400 semantic incongruity ERP paradigm to assess their knowledge of two groups of words: those expected to be known to the participants (high-frequency, familiar words) and those expected to be unknown (low-frequency, unfamiliar words). All three measures showed reliable differences between the known and unknown words. EM and pupillometry thus may provide insight into receptive vocabulary knowledge similar to that from ERPs. The development of additional implicit assessment techniques may increase the feasibility of receptive vocabulary testing across a wider range of participant groups and testing situations, and may make the conduct of such testing more accessible to a wider range of researchers, clinicians, and educators.

Keywords Eye movements · Pupillometry · Event-related potentials · Receptive vocabulary

One of the great challenges in the study of cognition and learning is to know what an individual knows. What would seem to be the most direct method—just asking them—is fraught with many limitations. For example, the representations in the cognitive architecture and the processes that operate on them are not always available to conscious access. Even to the degree that they are, it may be difficult for the typical adult participant to describe them using commonplace language. And asking is simply impossible for those with limited or absent verbal communication abilities, such as infants, small children, those with some kinds of developmental disabilities, and nonhuman animals.

For this reason, other methods to assess learning and cognition have been developed that rely instead on observations of the participants' behavior. Inferences are then drawn between these behavioral measures and the more elusive constructs of interest. One such behavioral measure is reaction time (RT): Given the assumption that cognitive processes unfold in time, the measurement of how long it takes an individual to respond to stimuli that vary along different dimensions can provide insight into the number of processes being engaged, the difficulty of those processes, or the time needed to access the stored representations, all of which might meaningfully differ across experimental conditions (Donders, 1969; Posner, 2005). Another example is the habituation paradigm, used to study cognition in infants, in which looking time is used as a measure of interest or stimulus novelty in babies: Babies will visually attend to a stimulus until they no

K. Ledoux · E. Coderre · L. Bosley · E. Buz · I. Gangopadhyay · B. Gordon
Cognitive Neurology/Neuropsychology Division, Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

B. Gordon
Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA

K. Ledoux (✉)
Cognitive Neurology/Neuropsychology, Johns Hopkins University, 1629 Thames St. Ste. 350, Baltimore, MD 21231, USA
e-mail: kledoux1@jhmi.edu

Published online: 11 March 2015

longer perceive it as novel, and they will look away when they tire of it. Making small alterations to a stimulus after the child has looked away from it allows researchers to determine, on re-presentation, whether the child is aware of the changes (if he or she spends time looking at it again) or not (if he or she doesn't; Colombo & Mitchell, 2009; Fantz, 1964). These and other behavioral methods have an extensive history of use in the fields of cognitive psychology, cognitive neuroscience, and education, and have contributed greatly to our current understanding of cognition and learning.

These methods are not without limitations themselves, however. One important limitation is that behavioral techniques often require an understanding of task instructions and/or the execution of complex behaviors in responding, making them difficult or impossible to use with certain participant populations. Another limitation is the ability to generalize their use to participant populations other than normal adults. For example, making inferences about age-related changes in cognitive processing from RT studies can be difficult, because such changes may be confounded with age-related changes in motor responses. Even the habituation technique described above, which has been used successfully to study cognition in infants, may present difficulties of interpretation for groups in whom looking behavior may be unreliable, such as low-functioning individuals with autism. Finally, many of these behavioral techniques depend on a participant's motivation to engage in and complete the task, something that again might vary tremendously across participant groups (and even across testing sessions, for individual participants).

For this reason, recent years have seen an emphasis on the development of assessment techniques that do not necessarily rely on an explicit behavioral response. These more implicit assessment methods (which include techniques such as functional magnetic resonance imaging, event-related potentials, eye movement monitoring, and pupillometry, among others) have the advantage of being useable even with individuals in whom more overt verbal or behavioral responses would be difficult or impossible to reliably obtain. Thus, they may prove especially useful in the study of cognition and learning in infants, small children, and patient populations, groups that frequently have been underrepresented in studies of cognition due to the difficulty of testing them.

Event-related potentials in the study of receptive vocabulary knowledge

An example of the development of an implicit technique to study an area of learning comes from the prolific recent use of event-related potentials (ERPs) to study receptive vocabulary knowledge. ERPs are event-locked changes in the scalp-recorded electroencephalogram (EEG). ERPs provide information with very fine temporal resolution (on the order of milliseconds) about how cognitive processes unfold in real time. Most importantly, ERPs can be meaningfully observed even in the absence of an overt behavioral response.

Separable ERP components have been reliably associated with separable aspects of cognitive processing; the evaluation of different ERP components can give a direct indication of the involvement of individual cognitive functions. One such component that has been reliably associated with cognitive operations is the N400, a negative-going deflection that peaks approximately 400 ms after the onset of a meaningful stimulus (such as a word or picture; Kutas & Hillyard, 1980). The N400 has been demonstrated to index semantic integration processing, by which the meaning of new stimuli is understood as being a part of, and integrated with, the current semantic context. Stimuli that are easier to integrate with their preceding context (for example, those that are semantically congruent with their context) elicit a smaller-amplitude N400 than words or pictures that are more difficult to integrate (for example, those that are semantically incongruent) with their context (Brown & Hagoort, 1993; Holcomb, 1993; Kutas & Federmeier, 2000; Van Petten & Kutas, 1991). The difference in the amplitude of the N400 between congruent and incongruent conditions has been called the N400 congruity effect.

This elicitation of the N400 congruity effect has been exploited to study receptive vocabulary knowledge by pairing a word with the meaningful context of a concurrently presented picture. A reduction of the amplitude of the N400 component is observed when the picture matches the named word (reflecting the ease of integrating the matching stimuli) relative to when it does not (reflecting the greater difficulty in integrating the incongruent stimuli). Critically, the integration of the auditory word with the picture context, and the resultant reduction of the N400 in cases of congruity, depends upon the listener/viewer having knowledge of the word's meaning. To the extent that the word is unknown to the participant, the integration between the word and its context cannot be eased by congruity because semantic knowledge cannot be brought to bear on the situation. In this way, an N400 congruity effect is predicted between spoken words and pictures, but *only* for words that are known to the participant. In the case of unknown words, no reduction of the amplitude of the N400 component would be expected, because the participant cannot use semantic knowledge to ease integration in the congruent condition. In other words, there should not be a difference in the ERPs for the congruent and incongruent conditions for unknown words because if the participant does not know the meaning of the word, he or she cannot assess whether the word and picture match.

A number of studies have confirmed these predictions. For example, Connolly and colleagues (Byrne et al., 1999; Connolly & D'Arcy, 2000; Connolly, D'Arcy, Newman, & Kemps, 2000; D'Arcy et al., 2003; Marchand, D'Arcy, & Connolly, 2002) used this type of N400 congruity paradigm

to assess receptive vocabulary in a series of studies with various participant groups, including healthy adults, typically developing children, adults with aphasia, and a child with cerebral palsy (for whom motor activity, and thus behavioral response, was limited). They used a variety of behavioral measures to estimate each participant's receptive vocabulary level, then presented congruent and incongruent pictures with words that were within or beyond that participant's vocabulary level. In each case, a larger N400 was observed to the auditory word in the incongruent condition than in the congruent condition—but only for words that were within the participant's vocabulary level. Other research has revealed similar results with a variety of participant groups (see, for example, Friedrich & Friederici, 2004, 2005a, b, 2010; Henderson, Baseler, Clarke, Watson, & Snowling, 2011; Torkildsen et al., 2008), and has even demonstrated the elicitation of the effect following training on new words (Friedrich & Friederici, 2008; Junge, Cutler, & Hagoort, 2012; Key, Molfese, & Ratajczak, 2006; Ojima, Nakamura, Matsuba-Kurita, Hoshino, & Hagiwara, 2010; Torkildsen et al., 2009). These findings thus support the utility of ERP measures to help discriminate sets of known words from sets of unknown words, and demonstrate the capability of this technique to be used in the testing of a wide variety of participant groups (including those who may otherwise have struggled to make overt behavioral responses).

Despite this demonstrated utility, the ERP paradigm carries some important potential limitations to the study of receptive vocabulary knowledge. First, this technique does not easily allow examination of the brain's response to a single item. Because the signal (the electrical activity of the brain) is relatively weak when compared to the multiple sources of noise in the EEG recording (such as eye movements, blinks, and muscle activity, all of which contribute to the recorded electrical signal), it is only through averaging the time-locked signal to multiple trials of a like type (or from the same experimental condition) that the brain potential can be sufficiently isolated from the noise. The greater the number of trials included in the average, the better the chance of eliminating more of the noise and observing more of the brain's activity. For this reason, studies that have used ERP paradigms to study receptive vocabulary have done so by comparing the averaged response across many trials of like words. When this is done, we see that the ERPs to known words differ from those to unknown words. This is very useful information, to the extent that researchers and experimenters can determine a priori pools of known and unknown words. Yet imagine the case of a clinician or a teacher, who wishes to use some objective measure to determine precisely which words are known and which are not, on a single word basis. This kind of determination from ERP data is not possible, because the observed signal to a single word on a single trial is simply too noisy. (One possible solution to this problem would be to run multiple trials of a single word and to average these together, although this approach also contains potential limitations, such as the fact that ERPs are also very sensitive to lexical repetition.) For this reason, the development of other implicit measurement techniques that provide benefits similar to those of ERPs, but that might also allow the examination of responses to single words, would be useful.

A second potential limitation to the use of an ERP paradigm to assess vocabulary is that ERP equipment (and the training necessary to learn to use it) is not necessarily readily accessible to the wide variety of professionals (clinicians, teachers, speech-language pathologists, etc.) and to the families who work with them, who might benefit from having a greater knowledge of an individual's receptive vocabulary level. ERP equipment has certainly become more affordable in recent years (even for very high-density electrode systems), and because of this and an increase in the application of ERPs to the study of cognition, more and more research groups in psychology, speech-language science, and other areas of cognitive neuroscience do have the potential to use this technology to study cognitive function. However, the cost remains relatively prohibitive for widespread access and use, especially compared to other technologies and methods of investigation.

Finally, some participant groups (such as infants, young children, and patients with acquired or developmental disorders) may prove less amenable to ERP testing, which may take longer than other methodologies (due to the need to acquire many trials) and which often involves the lengthy application of equipment that may prove intolerable to some (such as individuals with autism, who are frequently observed to dislike things being placed on their heads). Additionally, for reasons mentioned above, the isolation of the brain's electrical activity generally benefits from a reduction in electrical activity from other sources, such as muscles or the eyes. Keeping such sources of extraneous electrical noise (artifacts) out of the ERP is usually best accomplished through instructions to participants, for instance, to refrain from moving the body and the eyes and to refrain from blinking during critical portions of the trial. Participant groups who have more difficulty understanding such instruction or complying with them will necessarily have noisier data. Traditionally, the detection of artifacts during critical trials resulted in data loss, as such trials would be rejected from the final analysis. Recently, data-analytic procedures have been developed that allow for artifact correction (in place of artifact rejection), but even these methods cannot guarantee the removal of extraneous noise and often still result in data loss that may make the inclusion of participants from certain groups difficult or impossible.

Other implicit assessment techniques: Eye movement monitoring and pupillometry

For these reasons, in the present study we aimed to determine whether the assessment of receptive vocabulary knowledge could be extended to other implicit measurement techniques—specifically, eye movement monitoring (EM) and pupillometry. Both of these techniques have been used in the study of a variety of cognitive processes, as we discuss below. Additionally, they both may prove capable of providing more stable or reliable information about single words from single trials, which could be of tremendous benefit in determining which individual words are known to an individual participant. Furthermore, the equipment used to collect EM and pupillometry data is generally available at lower costs than ERP equipment, and may be available to some clinicians or instructors to whom ERP equipment is not. To the extent that EM and pupillometry data could be collected from paradigms that complement those used to study receptive vocabulary using ERPs, this might extend the application of implicit techniques generally to wider participant populations.

Eye movement monitoring EM and pupillometry both have long histories of application to the study of various aspects of cognitive processing. Eye movements have long been taken to reflect current cognitive operations (Just & Carpenter, 1976; Rayner, 1998), and thus have been used extensively to study various aspects of cognition, especially language. Although much of this application has been in the study of reading, eye movements have also played an important role in our understanding of the interface between spoken word recognition and aspects of language processing such as phonology or semantics. For example, Cooper (1974) demonstrated that participants would move their eyes to various pictures of objects as they heard those objects named in a story. More recently, similar demonstrations have come using the visual world paradigm (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Tanenhaus, 2007; Tanenhaus, Magnuson, Dahan, & Chambers, 2000; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, participants are asked to look at visual displays while listening to speech (which might include explicit instructions to manipulate the objects, either with their hands or with a computer mouse, or might be presented in a more passive way without instruction). These studies have consistently shown a tight time-locking between the unfolding of the auditory signal and participants' eye movements, such that participants will move their eyes to named objects in the display as soon as they can recognize even a minute fragment of the auditory stimulus. These types of paradigms have been used to study the time course of spoken word recognition, for example, by identifying the time at which two objects with names that share speech onsets become disambiguated, both in adults (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001) and in children (Fernald, Perfors, & Marchman, 2006; Fernald, Swingley, & Pinto, 2001; Swingley & Fernald, 2002; Trueswell, Sekerina, Hill, & Logrip, 1999). Recently, and importantly for our purposes, such paradigms have also been used in the study of semantic activation during speech perception (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005, 2011; Myung, Blumstein, & Sedivy, 2006; Myung et al., 2010; Yee, Huffstetler, & Thompson-Schill, 2011). For example, Yee and Sedivy (2006) showed that as participants heard a spoken word, their eyes were more likely to move not only to the named object in the display, but also to a semantically related object (for instance, if the spoken word was *lock*, eye movements were more likely to a picture of a lock and also to a picture of a key, relative to unrelated pictures in the display).

Odekar, Hallowell, Kruse, Moates, and Lee (2009) used a similar visual-world-type paradigm to determine whether patterns of eye movements could be used as valid and reliable indicators of semantic priming. Participants were presented with a printed prime word (such as *marriage*), followed by a display of three objects. On related trials, one of the three objects (e.g., a ring) shared a semantic/associative relationship with the prime, whereas the other two (for example, a nail and an ear) did not. Odekar and colleagues found that participants were faster to look at a given picture when presented in the related condition (i.e., preceded by a related prime word) as compared to when the same picture appeared in the unrelated condition (i.e., preceded by an unrelated prime word). Specifically, participants looked longer at related pictures both on average (mean fixation duration) and on initial processing measures (first fixation duration). They were also faster to move their eyes to a picture for the first time (latency to first fixation) in the related condition. Additionally, proportional measures, such as proportion of fixation durations on the target picture, showed that participants spent relatively greater amounts of total looking time over the course of the trial on pictures that were related to the preceding word. Finally, they found that a higher percentage of first fixations across trials were made to the named object on related trials, relative to unrelated trials. These results suggest that semantic priming is indeed reflected in differential patterns of eye movements in a visual-world-type paradigm. On the basis of these results, we hypothesized that the semantic relationship between an auditory word and its visual image would similarly affect eye movements, but only to the extent that that relationship was known to the participant. For both semantically related items and for known words, we expect to see an advantage in processing that is conferred by the knowledge of a semantic match between the auditory cue and the picture. Such

an advantage would not be conferred in the cases in which the items do not share a semantic relationship (either because they do are not close in the semantic network, or because one does not have the necessary knowledge to determine if they are related), and this should lead to greater relative difficulties in processing. In this way, we expected to see similar differences as those observed by Odekar and colleagues when comparing eye movements in our known and unknown word conditions.

Pupillometry *Pupillometry* measures changes in pupil dilation to study various aspects of cognitive processing. Pupillary dilation can be affected by many environmental or participant-internal events (such as changes in lighting, emotional arousal, or the onset of stress; Beatty & Lucero-Wagoner, 2000; Goldwater, 1972; Hess, 1965; Hess, Seltzer, & Shlien, 1965; Loewenfeld, 1993). Importantly, however, changes in pupil size are also observed in relation to the demands elicited by cognitive tasks, and these changes have been shown to occur independently of other influences (Goldwater, 1972; Karatekin, Marcus, & Couperus, 2007). Such changes are generally observed by time-locking changes in pupil diameter to the onset of stimuli that elicit various cognitive processes, and thus are often referred to as task-evoked pupillary reflexes (Beatty, 1982; Kahneman & Beatty, 1966; Kahneman, Beatty, & Pollack, 1967). Such task-specific changes in pupillary diameter have long been associated with attentional engagement and information processing: pupillary dilation has been shown to increase with task difficulty in many paradigms, and has thus been taken as a measure of resource recruitment (Beatty, 1982; Hess & Polt, 1964; Hoeks & Levelt, 1993; Kahneman & Beatty, 1966). Recently, Kuipers and Thierry (2011, 2013) examined pupillary responses recorded during an N400 picture–word congruity paradigm with high frequency words. For adults, congruent picture–word pairings elicited both a reduction in the amplitude of the N400 and smaller pupil sizes, relative to incongruent pairings. In a second study comparing monolingual and bilingual children, similar N400 congruity effects were observed in both groups. However, only the bilingual children showed the congruity effect in the pupillometry measures, suggesting that there may be developmental changes in resource recruitment during this task that might emerge earlier in children who speak more than one language.

For our purposes, we measured pupillary dilation during a four-alternative forced-choice task. We expected that participants would show greater cognitive resource recruitment when asked to select the picture that matched an unknown auditory word, relative to the very well-known and overlearned pairings between the known words and their visual depictions used in our study. We therefore anticipated that we would see larger changes in pupillary dilation in the unknown condition than in the known condition.

The present study

In the present study, we used all three of these implicit assessment techniques—EM, pupillometry, and ERP—to assess receptive vocabulary in a group of normal adult participants. The use of EM to study receptive vocabulary was novel to this study; two previous studies have used pupillometry for this purpose, but in a different paradigm than we employed here. We presented participants with two tasks involving pictures and auditory tokens of two sets of words: words that we expected would be known to all of the participants (such as *circle* and *dog*) and words that were expected to be unknown to most of the participants (such as *bilby* and *loquat*). In the *forced-choice task*, participants saw four pictures on the screen and heard one named; they were asked to use the mouse to select the named picture while EM and pupillary diameter data were recorded. In the *congruity task*, participants saw one picture on the screen and heard an auditory token that either matched (congruent condition) or did not match (incongruent condition) the picture. Participants indicated by button-press whether the picture matched or did not match the spoken word. ERP data were acquired during this task. In both tasks, half of the trials involved "known" words and half involved "unknown" words.

We made specific predictions for each of the three implicit measures. For EM, we predicted that eye movements would be faster and more reliable to the named picture for known words; for unknown words, looking behavior was expected to be more random and variable. Specifically, on the basis of the previous study conducted by Odekar and colleagues (2009) for semantic priming, we expected that measures of looking time (mean fixation duration, first fixation duration, and first dwell duration) would be longer to known items than unknown items. We also expected that participants would be faster to look at the named picture for the first time (latency to first fixation) and to return to the named picture once having moved the eyes away from it (latency to first refixation) for known items relative to unknown items. We expected that the proportional measures of time spent fixating the picture (proportion of fixation duration on stimulus) and of time spent looking at the named picture whether fixating or not (proportion of dwell time on stimulus) would both be greater for known words than for unknown words. Finally, we expected that the percentage of times that the named object would be the first and the last object fixated during a trial would be greater for known objects than for unknown objects. For pupillometry, we predicted that changes in pupillary dilation from baseline would be greater in the unknown condition, relative to the known, reflecting greater resource recruitment when the word's meaning was unknown. For ERPs, we expected to replicate previous studies that have used the N400 to assess receptive vocabulary knowledge by showing an N400 congruity effect for known words; that is, the amplitude of the

N400 component of the ERP was expected to be larger in the incongruent condition, relative to the congruent condition, but only for words that were known to the participant. For unknown words, prior knowledge could not be used to determine the congruity between the auditory word and the picture, and therefore the amplitude of the N400 was expected to be approximately the same for the congruent and incongruent conditions. Thus, no N400 congruity effect was predicted for the unknown condition.

Method

Participants

The participants were 23 adult, right-handed native speakers of English between the ages of 19 and 61 ($M = 35$ years; 70 % male, 30 % female). All had self-reported normal or corrected-to-normal vision and hearing. None of the participants reported cognitive, learning, or neurological impairment, and none were currently taking medication that might affect neurological or cognitive functioning. Participants were recruited from the Johns Hopkins University and surrounding community. The experimental procedures had been approved by the Johns Hopkins School of Medicine Institutional Review Board, and all participants gave written informed consent before participating in the experiment. All were monetarily compensated for participating.

Materials

Participants completed two separate tasks (a forced-choice task and a congruity task; see below) using the same set of 160 words and pictures. Eighty of the words (hereafter called "known words") were very high frequency (as assessed by the Subtlex US database; Brysbaert & New, 2009; $M = 3.14$, $SD = 0.6$), and were expected to be familiar to even very young children; these included words such as *cat*, *airplane*, and *camera*. The remaining 80 words ("unknown words") were low in frequency ($M = 0.85$, $SD = 0.5$), relatively unfamiliar words that were not expected to be known by the majority of the adult participants. Pretesting of these materials with a separate group of normal adults from the Johns Hopkins University community ($n = 15$) confirmed that these words were relatively unknown to this group. Examples of words in this set included *cherimoya*, *agouti*, and *cainito*. All words were highly imageable. High-resolution, color digital pictures were selected to represent each word. Pretesting with a separate group of normal adult participants ($n = 3$) ensured the suitability of the images as representations of their corresponding concepts. High-quality, digital auditory recordings of a female speaker pronouncing the name of each of the items were made using Audacity software, and were edited using Computerized Speech Lab Model 4150 software (KayPENTAX). The edited tokens ranged in length from 500 to 1,200 ms. The recordings were transcribed by two speech-language pathology graduate students naïve to the purposes of the experiment. Their transcriptions were compared to that of a speech-language pathologist who assisted with the recordings to ensure that the words were completely audible. Volume was normalized across the tokens.

Procedure

Participants completed two testing sessions. In one, participants completed either the forced-choice task (during which EM and pupillometry data were collected) or the congruity task (during which ERPs were recorded), along with the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007). In the second session, participants completed the second of the two tasks (forced choice or congruity), along with a word familiarity rating task. We chose to collect EM and PD data during the forced-choice task, and ERP data separately during the congruity task, to maintain consistency with previous studies that have used similar paradigms to examine semantic processing (as described in the introduction). The three tasks are described in further detail below.

Forced-choice task In the forced-choice recognition task, presented in E-Prime (version 2.0.8.74), participants were asked to use the mouse to select one of four pictures presented simultaneously on a computer screen after having heard one of the objects named. On each trial, participants saw a fixation cross at the center of the screen for 1,000 ms. The fixation cross remained on the screen as the four pictures appeared, one in each corner of the screen (see Fig. 1). Twenty milliseconds later, participants heard one of the pictures named. On all trials, the three distractor items were drawn from the same knowledge category (known/unknown) to prevent participants from using a process of elimination on unknown trials. The pictures remained on the screen until the participant selected one with a mouse click, or for a maximum of 5 s. There were 160 trials, one per experimental item. These were presented in eight blocks of 20 trials, in which half were known targets and half were unknown (pseudorandomized within blocks). One practice trial served to familiarize participants with the paradigm at the start of the experiment. We simultaneously collected eye movement and pupillometry data using an ASL Model 504 eyetracking system. The pupil diameter was measured horizontally and was recorded every 17 ms in pixels. Reaction times and accuracy were also recorded.

Congruity task In the congruity paradigm, also presented in E-Prime, a picture was presented on the computer screen, followed immediately by the auditory presentation of a single word, which either matched (congruent condition) or did not

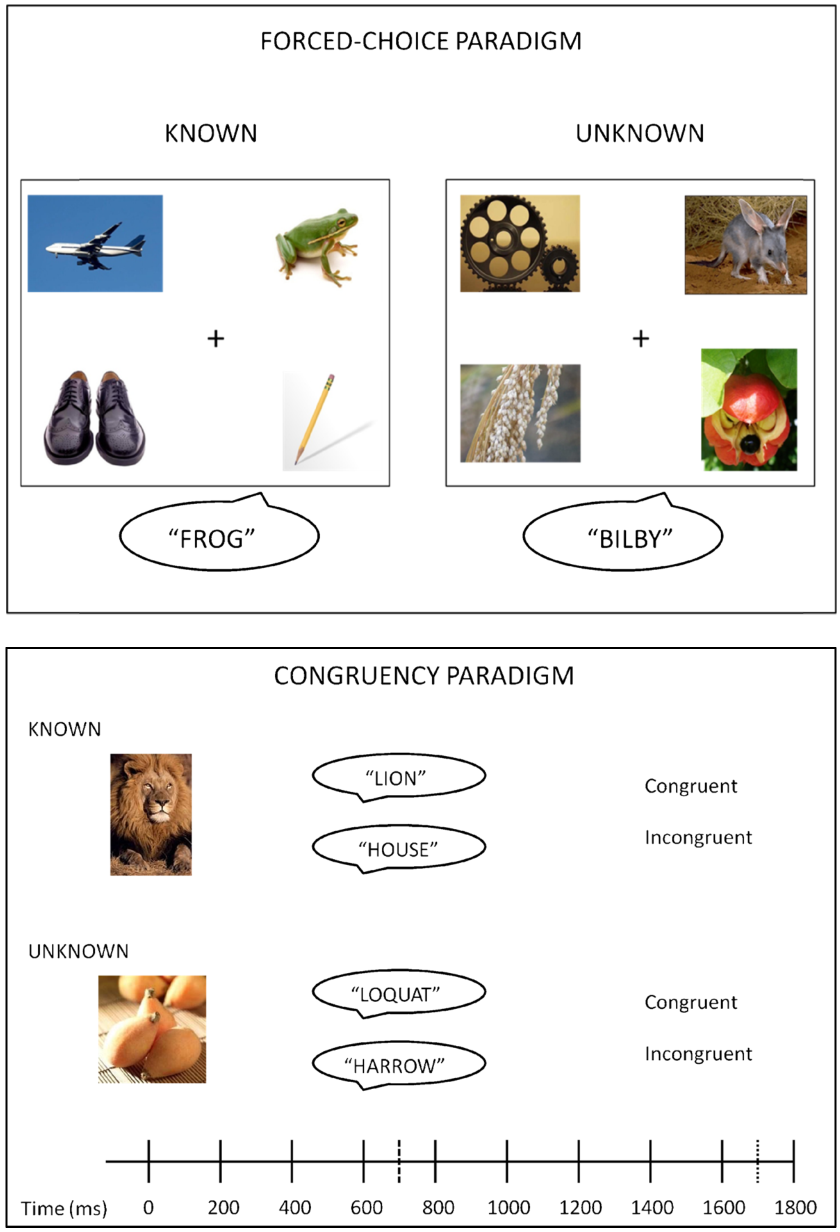


Fig. 1 Schematic representations of the forced-choice paradigm (top) and the congruency paradigm (bottom). See the text for details of each. For the unknown condition in the forced-choice example shown above, the pictured objects are (clockwise, from top left) *pinion*, *bilby*, *ackee*, and *millet*. For the unknown condition in the congruency example above, the pictured items are *loquats*

match (incongruent condition) the pictured item (see Fig. 1). As in the forced-choice task, the mismatching pictures for the incongruent condition were chosen from the same knowledge category (known/unknown) as the auditory token, to avoid strategic responses based on a process of elimination. Additionally, care was taken to ensure that the incongruent word–picture pairs did not share the same initial phoneme. A red fixation point (presented for 1,000 ms) began each trial, followed by the presentation of the picture for 700 ms. The auditory token was then presented (varying in length from 500 to 1,200 ms). The picture remained on the screen throughout the duration of the auditory token and for another 1 s after its offset, during which time responses were prohibited. A response screen (indicated by a green fixation point) was then presented for up to 5 s or until participants made a response. Participants used two buttons on a button box to indicate whether the auditory word and the picture matched or did not match. They were also instructed to keep their eyes fixated on the center of the screen, to move as little as possible, and to refrain from blinking during the presentation of the picture and the auditory token. This was done to minimize artifacts in the EEG signal. There were 320 trials, two per experimental item (one congruent pairing, one incongruent pairing). These were presented in eight blocks of 40 trials each, in which ten trials of each type (known–congruent, known–incongruent, unknown–congruent, and unknown–incongruent) were pseudorandomized. In an initial trial block, ten nonexperimental items were presented to familiarize participants with the

task. High-density ERPs were recorded during the congruity task at 250 Hz using a Geodesics 256-channel sensor net (see Fig. 2) and NetStation version 4.3. Impedances were kept under 50 kΩ, where possible.

Word familiarity task Participants were asked to participate in a word familiarity posttest after the EM, pupillometry, and ERP testing had been completed. In the posttest, the participants were presented with each of the 160 auditory tokens and asked to rate their familiarity with the word on a scale from 1 (*very unfamiliar*) to 9 (*extremely familiar*), with an additional option of 0 (*no familiarity whatsoever*).

Data processing and analysis

The data from each of the three implicit measures were processed and analyzed separately. For all analyses, effect sizes are reported as Cohen's d (for t tests) or eta-squared (η^2, for analyses of variance [ANOVAs]), calculated for the within-subjects measures.

Eye movements Fixation data were analyzed using ASL Results (Applied Science Laboratories, 2009). Each presentation slide was divided into five areas of interest: the four picture stimuli and the fixation cross in the middle of the screen. After discarding any trials in which more than 50 % of the trial was not detected, an average of 78 % of unknown and 75 % of known trials remained for statistical analysis.

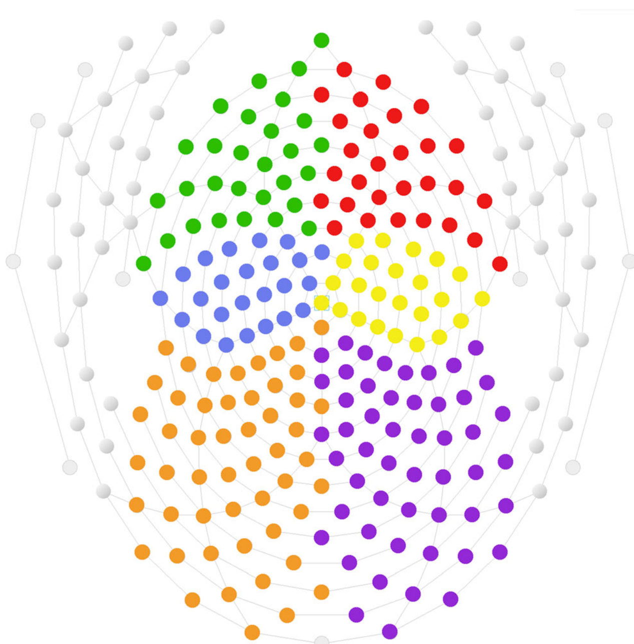


Fig. 2 The 256-channel electrode montage. Color is used to indicate the six electrode groupings used in the analysis: Green and red indicate frontal electrodes on the left and right, respectively; blue and yellow indicate central electrodes; and orange and purple indicate parietal electrodes

For each trial, we calculated a number of dependent variables derived from fixation and dwell time measures. Fixations were defined as periods of time (in milliseconds) for which eye gaze remained at a specific location on the screen. Fixation onsets were defined as a stable gaze duration of at least 100 ms and a visual angle variation of less than 1 degree. Fixation offsets were defined as three or more sequential samples that deviated from the fixation start location by more than 1 deg of visual angle. Dwell time was defined as the amount of time (in milliseconds) spent looking at the named picture, with or without fixation (i.e., the time that the eyes were in the region of interest of the named picture, whether they stayed in one spot long enough to reach the threshold for fixation). From these measures, the following dependent variables were derived:

Total number of fixations: the number of fixations made during the entire trial
Mean fixation duration: the average length of all fixations within the area of the named picture
First fixation duration: the length of the first fixation within the area of interest for the named picture
First dwell on stimulus: total time spent looking at the named picture during the first dwell
Latency to first fixation: the amount of time that passed before the first fixation on the named picture
Latency to first refixation: the amount of time that passed before a refixation occurred in the area of interest of the named picture (i.e., the amount of time to come back to fixate on the named picture after the eyes had left the region of this picture)
Proportion of fixation duration on the stimulus: the proportion of fixation duration time on the named picture relative to total fixation duration time (i.e., fixation duration on the named stimulus/total fixation durations for all four pictures)
Proportion of dwell time on stimulus: the proportion of the trial that was spent looking at the named picture, with or without fixation (i.e., dwell time on the named stimulus/length of the trial)
Percentage of trials first fixated: the percentage of trials (out of all good trials) on which the named object was the first picture to be fixated
Percentage of trials last fixated: the percentage of trials (out of all good trials) on which the named object was the last picture to be fixated

Pupillometry To convert the pixel measurement to millimeters, a scaling factor was calculated using a model eye provided by ASL to simulate the image received from a real eye. When viewed by the eye tracker optics, the model eye simulates a pupil image and corneal reflection. To calibrate the

pupil diameter, we positioned the model eye so that the white circle was at a normal eye distance from the optics, oriented so that the corneal reflection appeared below the pupil.

We used the Interface software to discriminate on the model pupil image, recorded the pupil diameter value on the computer screen digital display window, and computed the scale factor. To compute a scale factor (millimeters per eyetracker unit), the diameter of the white circle (4 mm) was divided by this value. To perform the analyses, the recorded pupil diameter values were converted to millimeters by applying this scale factor (value in millimeters = scale factor × recorded value).

Pupillary responses were analyzed using Microsoft Office Excel 2007 and IBM SPSS Statistics 19. Prior to the statistical analyses, the data were cleaned by removing artifacts due to excessive blinking and by replacing small blinks by linear interpolation. Trials in which 20 or more data points in a row (340 ms or more) were missing due to a lack of fixations were discarded. After discarding the bad trials, an average of 75 % of the unknown and 81 % of the known trials remained for statistical analysis. For each trial, the average pupil diameter during the 200 ms preceding the stimulus onset was subtracted from the task-evoked pupil diameter. Pupil diameters were then converted to millimeters by applying the scale factor. The data were expressed as millimeter deviations from the pretrial baseline. We calculated three dependent variables from the pupillometry data:

Peak dilation: the size of the largest absolute change in pupil size from baseline
Mean change in pupil size: average change in pupil size from baseline across the trial
Maximum percent change in pupillary dilation: the proportion of the peak change in pupil size to baseline pupil size

ERPs The EEG data were preprocessed using EEGlab version 10.2.2 and MATLAB version 8.1. The data were first filtered using a 0.1- to 30-Hz bandpass filter and referenced using an average reference transform to the Cz electrode. Correction for eye movement artifacts was performed by first running a principal component analysis (PCA) on each participant to identify the number of components required to explain 99 % of the data. Independent component analysis (ICA) was then performed using the specified number of components. Following ICA decomposition, eye movements, blinks and other noise components were manually identified and removed from the data.

The resulting cleaned continuous data was segmented into epochs time-locked to the onset of the picture stimulus. Segments extended from 800 ms before to 1,000 ms after the auditory stimulus, in order to include the full response to the picture (presented at –700 ms). Additional bad epochs were identified and rejected using a joint probability computation. The resulting segments were baseline-corrected using data from the first 100 ms of the segment. In total, an average of 97 % of unknown and 97 % of known trials were included in the statistical analysis.

For the purposes of statistical analysis, the electrodes were broken into six topographic regions across the scalp, including left and right clusters for the frontal, central, and parietal regions (see Fig. 2). The data from these clusters were collapsed over all electrodes. An N400 window of interest was determined on the basis of latency expectations derived from the literature, visual inspection of the waveforms, and running t tests. For running t tests, the raw data were collapsed into 24-ms bins with 12 ms overlap, and the average amplitudes were compared between congruencies within each bin by using paired-sample t tests. More than five bins in a row showing significant differences between conditions ($p < .05$) was deemed a significant window. On the basis of these methods, N400 congruency effects were examined across the window of 450–700 ms post-auditory-onset. We first performed a 2 (knowledge: known, unknown) × 2 (congruency: congruent, incongruent) × 3 (site: frontal, central, parietal) × 2 (hemisphere: left, right) overall ANOVA for the window extending from 450 to 700 ms after sound presentation; significant main effects or interactions were then followed up with additional ANOVAs and post-hoc t tests. For the sake of clarity, we only report and follow up significant ($p < .05$) main effects or interactions.

Results

Behavioral data

Forced-choice task Across all trials, participants selected the correct (named) picture on 79.7 % of trials. Participants had significantly higher accuracy for known trials (99.9 %) than for unknown trials (59.5 %), $t(22) = 16.07$, $p < .0001$, $d = 3.35$. Participants took an average of 1974.8 ms to make their selection across all trials. They were significantly faster to respond on known trials ($M = 1{,}370.4$ ms) than on unknown trials ($M = 2{,}579.3$ ms), $t(18) = 17.76$, $p < .0001$, $d = 4.07$.

Congruity task Across all trials, participants made a correct response on auditory/picture congruity on 65.7 % of trials. To compare accuracy on the congruity task, a 2 (congruency: congruent/incongruent) × 2 (knowledge: known/unknown) repeated measures ANOVA was run on the mean accuracy for each participant. There was a significant interaction of congruency and knowledge [$F(1, 22) = 72.51$, $p < .0001$, $\eta^2 = .74$]. Post-hoc paired-samples t tests showed significantly higher accuracies for unknown incongruent (86.4 %) than unknown congruent trials (33.2 %), $t(22) = 8.54$, $p < .0001$, $d = 1.78$; this pattern may reflect a bias on unknown trials toward responding that the auditory cue did not match the picture. We found no accuracy differences between known

congruent (98.2 %) and known incongruent trials (98.8 %), $t(22) = 1.53$, $p = .14$, $d = 0.32$.

To compare RTs on the congruity task, a 2 (congruency: congruent/incongruent) × 2 (knowledge: known/unknown) repeated measures ANOVA was run on the mean RTs. There was a significant interaction of knowledge and congruency [$F(1, 22) = 7.21$, $p < .05$, $\eta^2 = .002$]. Post-hoc paired-samples *t* tests showed significantly slower RTs for unknown congruent trials ($M = 927.8$ ms) than for unknown incongruent trials ($M = 878.7$ ms), $t(22) = 2.69$, $p < .05$, $d = 0.56$. No differences in RTs was apparent between known congruent trials ($M = 759.0$ ms) and known incongruent trials ($M = 793.0$ ms), $t(22) = 1.53$, $p = .14$, $d = 0.32$. That participants responded faster to congruent trials for known words, but to incongruent trials for unknown words, may again reflect a response bias in the incongruent condition.

Word familiarity ratings Known words were given significantly higher word familiarity ratings ($M = 8.99$) than were unknown words ($M = 2.58$), $t(22.05) = 31.35$, $p < .0001$, $d = 6.54$.

EM data

Table 1 presents the mean values on all of the dependent measures derived from eye movements for the known and unknown conditions. Sample eye movement data are shown in Fig. 3. We observed a greater total number of fixations for unknown than for known trials, $t(22) = 17.71$, $p < .0001$, $d = 3.69$. On average, mean fixation durations on the named picture were longer in the known than in the unknown condition, $t(22) = 2.39$, $p < .05$, $d = 0.50$. The length of the first fixation on the named picture was longer for known than for unknown trials, $t(22) = 3.68$, $p < .01$, $d = 0.77$. The length of the first dwell on the named picture was also longer for the known than for the unknown condition, $t(22) = 3.86$, $p < .001$, $d = 0.80$. The latencies to first fixation on the named picture, $t(22) = 8.56$, $p < .0001$, $d = 1.79$, and to refixation, $t(22) = 8.64$, $p < .0001$, $d = 1.80$, were both shorter for known than for unknown trials. The proportion of time spent fixating on the stimulus (i.e., proportion of fixation duration on the stimulus) was greater in the known than in the unknown condition, $t(22) = 12.99$, $p < .0001$, $d = 2.71$. The proportion of time spent dwelling on the stimulus (i.e., looking at the named picture, with or without fixation) was also greater in the known than in the unknown condition, $t(22) = 11.13$, $p < .0001$, $d = 2.32$. The stimulus was the first picture to be fixated on a significantly higher percentage of known than of unknown trials, $t(22) = 2.46$, $p < .05$, $d = 0.51$. Finally, the stimulus was also the last picture to be fixated on a significantly higher percentage of known than of unknown trials, $t(22) = 19.55$, $p < .0001$, $d = 4.08$.

Table 1 Means and standard deviations for each dependent variable in the eye movement monitoring and pupillometry data

Dependent Variable	Known		Unknown	
	M	*SD*	*M*	*SD*
Eye Movements				
Total number of fixations in trial	3.54	0.86	6.98	1.44
Mean fixation duration (ms)	416.4	146.8	354.7	147.5
First fixation duration (ms)	406.8	126.4	328.4	64.7
First dwell on stimulus (ms)	605.4	236.5	445.1	84.5
Latency to first fixation (ms)	742.7	103.2	1,045.0	221.2
Latency to first refixation (ms)	894.7	313.2	1,573.4	594.6
Proportion of fixation duration on stimulus (%)	63.87	17.00	34.14	8.99
Proportion of dwell time on stimulus (%)	42.14	9.61	24.03	4.73
Percentage of trials first fixated (%)	33.50	13.32	27.58	8.38
Percentage of trials last fixated (%)	90.76	8.13	44.98	7.55
Pupillary Dilation				
Peak dilation (mm)	5.43	1.51	7.53	2.30
Mean change in pupil size (mm)	0.01	0.70	1.31	0.74
Max percent change in pupillary dilation (%)	15.78	3.82	22.10	5.45

Pupillometry data

The pupillometry data are also summarized in Table 1. Larger peak dilations (relative to baseline) were observed for unknown than for known trials, $t(22) = 9.24$, $p < .0001$, $d = 1.93$. Mean changes in pupil size from baseline were also larger for unknown than for known trials, $t(22) = 8.32$, $p < .0001$, $d = 1.73$. The maximum percent change in pupillary dilation was also larger for unknown than for known trials, $t(22) = 10.86$, $p < .0001$, $d = 2.26$.

ERP data

ERPs for the four conditions, as well as topographical plots of the incongruent–congruent difference for the known and unknown conditions, are presented in Fig. 4.

A 2 (congruency: congruent/incongruent) × 2 (knowledge status: known/unknown) × 3 (site: frontal/central/parietal) × 2 (hemisphere: left/right) repeated measures ANOVA was performed on the average amplitudes over a window from 450 to 700 ms after sound presentation (shaded regions in Fig. 4). The full results can be found in Table 2. We observed a significant three-way interaction of knowledge, congruency, and site [$F(2, 44) = 6.14$, $p < .01$, $\eta^2 = .07$].

To explore this interaction, we performed a 2 (congruency) × 2 (knowledge) ANOVA for each site (collapsed over hemisphere). There was a significant interaction of congruency and

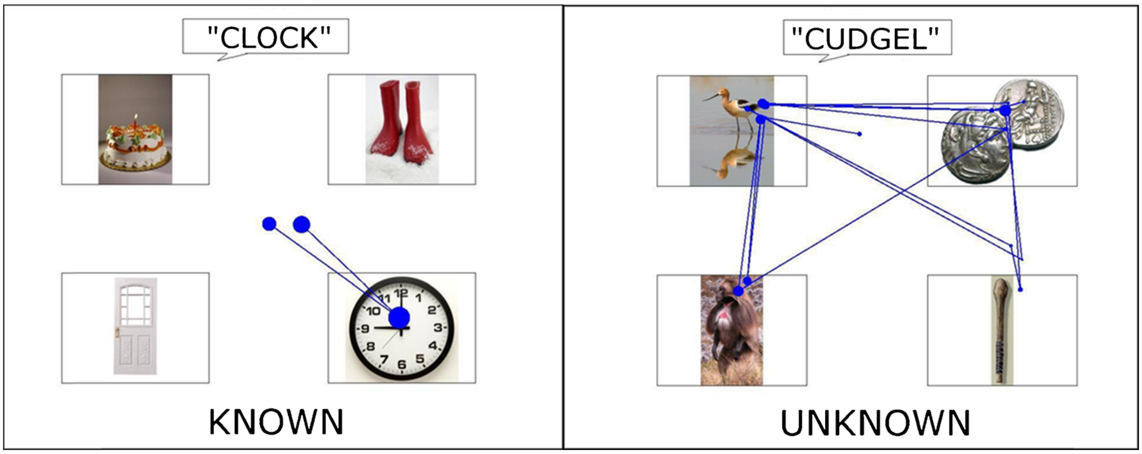


Fig. 3 Sample eye movement patterns for a single participant for a single trial in the known (left) and unknown (right) conditions. Blue dots indicate fixations (with the size of the dot indicating the length of the fixation)

knowledge at parietal sites [$F(1, 22) = 7.34$, $p < .05$, $\eta^2 = .02$]. We investigated this interaction at parietal sites by performing a two-way (congruency) ANOVA separately for the known and unknown words. For the known condition, a main effect of congruency emerged [$F(1, 22) = 7.69$, $p < .05$, $\eta^2 = .07$]: The mean amplitude observed to known congruent items (M = 0.08 μV, SE = 0.07) was more positive than that observed to known incongruent items (M = –0.21 μV, SE = 0.10). No such difference by congruency was observed for unknown items ($F < 1$, $p = .39$, $\eta^2 = .002$).

To summarize, a significant N400 congruency effect (a reduction in the amplitude of the N400 for congruent trials, relative to incongruent trials) occurred from 450 to 700 ms over bilateral parietal electrode locations—but only for the known items. No such N400 congruency effect was found for the unknown items.

Effects of familiarity observed to the picture In addition to the expected N400 congruency effect, we also observed an earlier difference in the waveforms recorded to the picture, before the auditory stimulus was presented. Because the auditory word had not yet been presented, any difference observed in this time window would be tied to knowledge differences for the pictures themselves, and could not be linked to congruity (since this was determined by the auditory stimulus). To examine this difference further, we collapsed the ERPs across congruence conditions to look at the differences elicited to the pictures in the known and unknown conditions; these ERPs are shown in Fig. 5.

Running t tests identified a sustained difference between the known and unknown conditions beginning approximately 200 ms after picture onset (i.e., –600 ms, relative to the onset of the auditory stimulus). The length of this significant window differed over sites. To compare these effects statistically, a 2 (knowledge: known/unknown) × 3 (site: frontal/central/parietal) × 2 (hemisphere: left/right) repeated measures ANOVA was run on the mean amplitudes for the known and unknown conditions (collapsed over congruency) over a window from 200 to 500 ms after picture presentation (–500 to –200 ms, relative to onset of the auditory token). This window was chosen as the minimum length at which all sites showed differences in the running t tests (Fig. 5). The ANOVA showed an interaction of knowledge and site [$F(2, 44) = 28.41$, $p < .0001$, $\eta^2 = .07$; see Table 3 for the full results]. To follow up this interaction, we collapsed over hemispheres and performed a two-way (knowledge) ANOVA for each site. Frontal sites showed a main effect of knowledge [$F(1, 22) = 30.00$, $p < .0001$, $\eta^2 = .01$], such that the mean amplitude to the known condition (M = –0.93 μV, SE = 0.16) was more positive than that to the unknown condition (M = –1.19 μV, SE = 0.18). This effect was also evident over central sites, where there was also a main effect of knowledge [$F(1, 22) = 11.91$, $p < .01$, $\eta^2 = .01$] due to a greater relative positivity to the known (M = –0.24 μV, SE = 0.10) than to the unknown (M = –0.39 μV, SE = 0.12) condition. At parietal sites, we also found a main effect of knowledge [$F(1, 22) = 31.89$, $p < .0001$, $\eta^2 = .02$], but here, the polarity of the effect was reversed: There was a greater relative positivity to unknown (M = 1.20 μV, SE = 0.14) than to known (M = 0.97 μV, SE = 0.13) items.

Thus, we observed differences in the response to the visual stimulus between the known and unknown conditions prior to the presentation of the (congruent or incongruent) auditory stimulus. At frontal and central electrode locations, this difference was in the form of a relatively more positive mean amplitude to pictures in the known condition, whereas at parietal sites, there was a greater relative positivity to pictures in the unknown condition. This difference onset early across all scalp locations (beginning approximately 200 ms after presentation of the picture) and extended in time for several hundred milliseconds (especially at parietal sites).

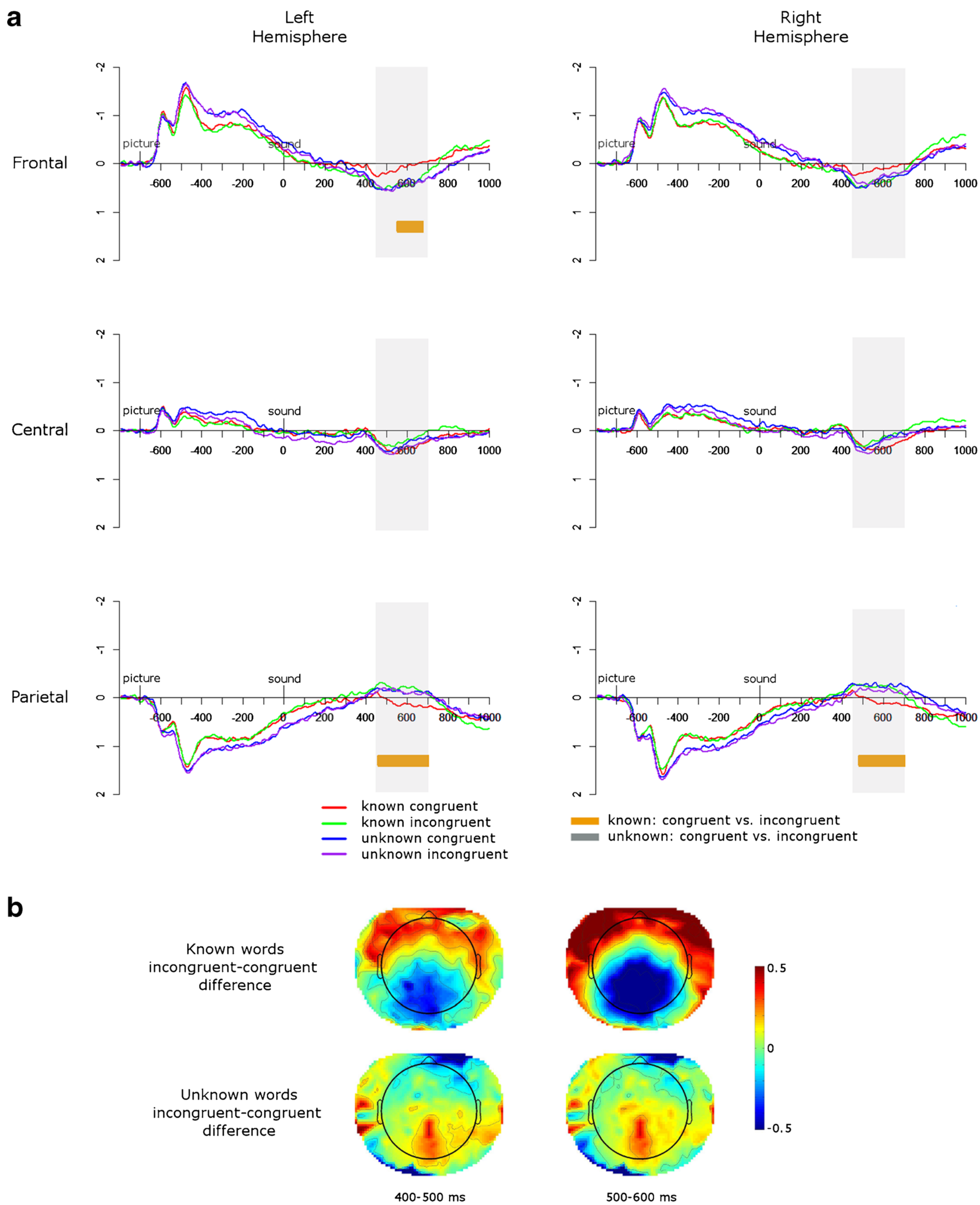


Fig. 4 (**a**) N400 effects of picture–word congruency for known and unknown words. ERPs are grand averages across all participants, collapsed across electrodes in the frontal, central, and parietal regions. Gray shading indicates the N400 window (450–700 ms after the onset of the auditory stimulus). (**b**) Topographic distribution of the difference waves (incongruent – congruent) for the known and unknown conditions across two time windows

Correlations among measures

To consider the relationship between knowledge status and our various dependent measures, we ran Pearson's correlations between several variable pairs separately for known and unknown items.

Behavioral measures with implicit measures First, we examined the correlations between the three behavioral measures and the implicit measures; these results are shown in Table 4. The first behavioral measure was the PPVT score for each participant. For known items, we observed several negative correlations between PPVT score and the EM duration

Table 2 Results from the 2 (knowledge) × 3 (congruency) × 3 (site) × 2 (hemisphere) analysis of variance on the average amplitudes over a window from 450 to 700 ms after sound presentation

Main effect or interaction	*F* Value	*df*	*p* Value	η^2 Value
Knowledge	<1	1, 22	.57	.002
Congruency	<1	1, 22	.85	.0002
Site	3.37	2, 44	<.05*	.11
Hemisphere	<1	1, 22	.43	.01
Knowledge× Congruency	1.30	1, 22	.27	.008
Knowledge × Site	2.31	2, 44	.10§	.03
Congruency × Site	3.89	2, 44	<.05*	.04
Knowledge × Hemisphere	2.50	1, 22	.13	.004
Congruency × Hemisphere	<1	1, 22	.98	<.0001
Site × Hemisphere	<1	2, 44	.96	.0002
Knowledge × Congruency × Site	6.14	2, 44	<.01**	.07
Knowledge × Congruency × Hemisphere	<1	1, 22	.78	.0002
Knowledge × Site × Hemisphere	1.52	2, 44	.23	.001
Congruency × Site × Hemisphere	3.95	2, 44	<.05*	.005
Knowledge × Congruency × Site × Hemisphere	<1	2, 44	.79	.0003

Significant effects are indicated: § trend, $p < .10$, * $p < .05$, ** $p < .01$

measures (such as mean fixation duration, first fixation duration, and proportion of dwell time on the stimulus), suggesting that larger vocabulary scores were associated with shorter looking times for known items. For unknown words, on the other hand, such correlations were not observed; the only significant correlation for this set was a positive correlation between PPVT score and the percentage of trials on which the named item was the last to be fixated.

The second behavioral measure was the RT on the forced-choice task. We ran correlations between these RTs and the EM and pupillometry measures, which were collected using the same paradigm. These are also shown in Table 4. Of note, for known words, we observed several positive correlations between the RT and EM measures, suggesting that longer times to select the named picture from the display were accompanied by longer looking times. This was not seen for unknown items, for which we saw only one marginally significant correlation between RT and the EM measure of first dwell time. There was, however, a significant negative correlation between RT and the mean change in pupil size for unknown items, suggesting that faster RTs were accompanied by smaller changes in pupil size for unknown items.

The third behavioral measure was RT on the congruity task, which we correlated with the N400 effect size from the concurrent ERP task. For the congruity task, RT effect sizes were calculated by subtracting the mean RT in the congruent condition from the mean RT in the incongruent condition, separately for known and unknown items. From the ERP data, the N400 effect was calculated by first calculating the difference wave (incongruent minus congruent) for each individual word, then finding the most negative peak in the difference wave within a window from 200 to 800 ms after sound presentation. The average difference wave amplitude within a 200-ms window around the most negative difference wave peak was then calculated, yielding an *N400 effect* measure for each individual word, which was averaged over known and unknown trials. As can be seen in Table 4, no significant correlations emerged between the RT effect size on the congruity task and N400 effect size.

Correlations among implicit measures We also ran correlations among the various implicit measures themselves. The results of these correlations are shown in Tables 5 (for known items) and 6 (for unknown items). Some patterns are worth highlighting. First, for both known and unknown words, the EM measures are all highly intercorrelated, as are two of the three pupillometry measures (peak dilation and maximum percent change in pupil dilation), suggesting that these measures may be tapping into the same underlying processes. There are also a number of significant correlations between the measures from the different implicit assessment techniques; for example, for known words (but not for unknown words), we observed significant correlations between the N400 effect size and EM measures such as the mean fixation duration and first fixation duration.

Discussion

In the present study, we used measures from three different implicit assessment techniques (eye movement monitoring, pupillometry, and event-related potentials) to study receptive vocabulary knowledge in normal adults. Specifically, we looked for differences between these measures to high-frequency, highly familiar words, which were expected to be known by all adult participants, and to low-frequency, unfamiliar words, which were expected to be unknown. The behavioral measures that we administered supported the distinction between these two sets of words. First, offline word familiarity ratings suggested that participants were very familiar with the high-frequency words in our known condition, and were relatively unfamiliar with the low-frequency words in the unknown condition. Additionally, on both the forced-choice and congruity tasks, participants were more accurate and faster when making responses to known than to unknown items. These results support the distinction between the two sets of words and suggest that these are appropriate sets in which to look for processing differences using our implicit measures.

The ERP technique has previously proven useful in detecting differences in receptive vocabulary knowledge for known

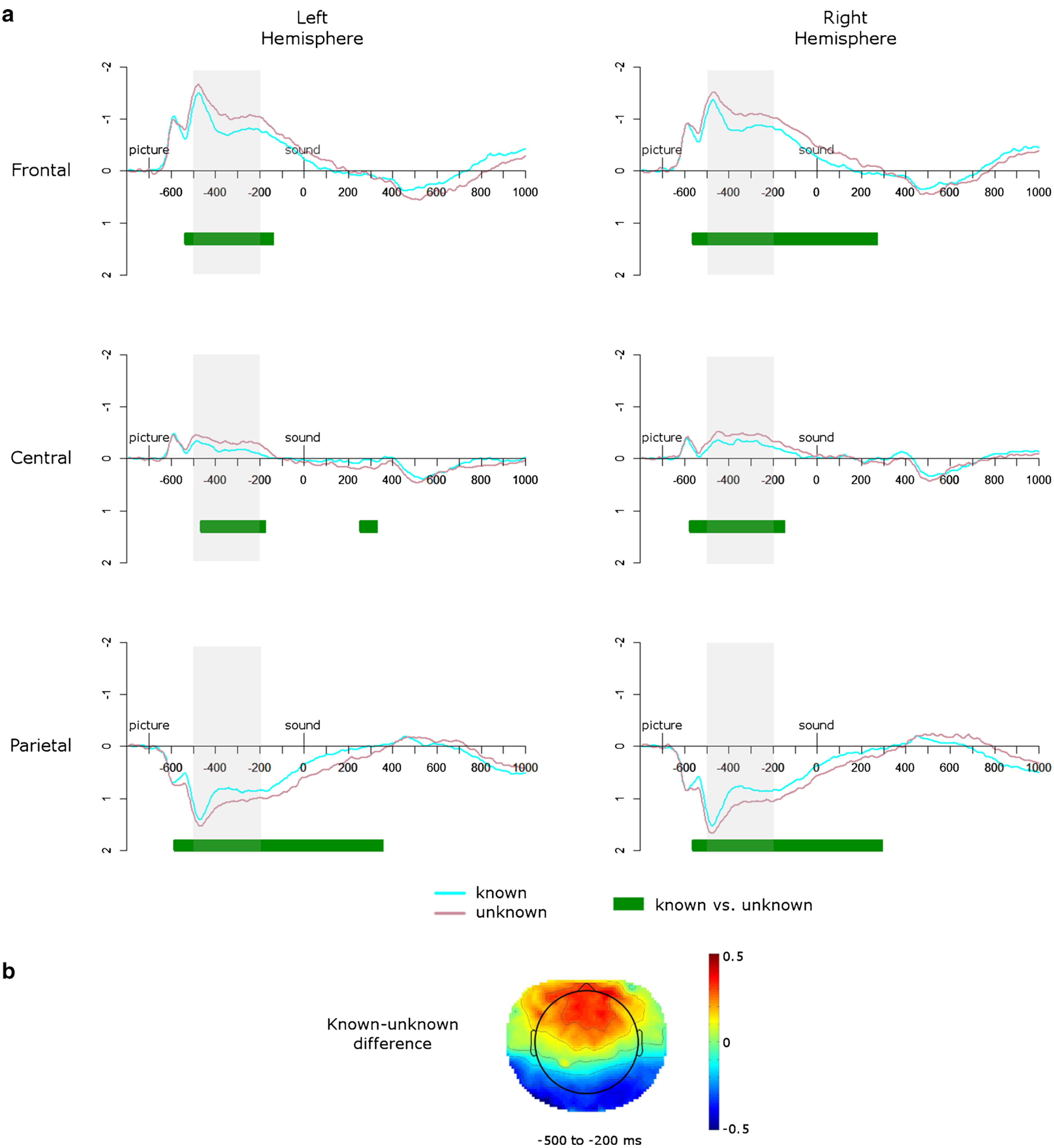


Fig. 5 (**a**) Effects of picture familiarity for known and unknown words. ERPs are averaged over congruent and incongruent conditions for known and unknown stimuli, and collapsed across electrodes in the frontal, central, and parietal regions. (**b**) Topographic distribution of the difference waves (known – unknown) across the time window of interest (–500 to –200 ms before sound presentation: gray shading)

and unknown words in a variety of participant groups. Our results with ERPs replicated those of several other research groups using a similar congruency paradigm (Byrne et al., 1999; Connolly & D'Arcy, 2000; Connolly et al., 2000; D'Arcy et al., 2003; Friedrich & Friederici, 2004, 2005a, b, 2010; Henderson et al., 2011; Marchand et al., 2002; Torkildsen et al., 2008). A reliable reduction in the amplitude of the N400 was observed in congruent versus incongruent word/picture pairings, but only for the items that were expected to be known to the participants. For the unknown word

Table 3 Results from the 2 (knowledge) × 3 (site) × 2 (hemisphere) analysis of variance on the known and unknown average amplitudes (collapsed over congruency) over a window from –500 to –200 ms before sound presentation

Main Effect or Interaction	*F* Value	*df*	*p* Value	η^2 Value
Knowledge	9.53	1, 22	$<.01^{**}$	.006
Site	25.7	2, 44	$<.0001^{***}$	.49
Hemisphere	<1	1, 22	.79	.0007
Knowledge × Site	28.41	2, 44	$<.0001^{***}$	.07
Knowledge × Hemisphere	<1	1, 22	.72	.0001
Site × Hemisphere	1.49	2, 44	.24	.006
Knowledge × Site × Hemisphere	<1	2, 44	.72	<.0001

Significant effects are indicated: $^{**}p < .01$, $^{***}p < .001$

pairings, no difference in the amplitude of the N400 was observed. The observed congruency effect for the known words would be expected if participants are able to use their knowledge of word meanings to use the picture as context when deciding whether the accompanying auditory token is congruent or incongruent. However, because such underlying semantic knowledge is not available in the case of the unknown words, the ease of processing that results from a congruency between a word and its context (and the resultant reduction of the N400) cannot occur. The N400 effect observed in our experiment was slightly delayed in latency relative to canonical N400 effects; this is likely due to the fact that we time-locked our ERPs to the onset of the auditory word. Processing may not have fully engaged until the offset of this word, or at least until enough of the word had been presented to allow for lexical selection.

In addition to the replication of the congruency effect for known (but not for unknown) words, we observed differences in the ERPs elicited by the visual stimulus itself, prior to the presentation of the auditory stimulus that determined congruency. Differences (in the form of a greater relative positivity to known pictures at frontal and central sites, and a greater relative positivity to unknown pictures at parietal electrode sites) were observed shortly after the onset of the picture stimulus, and remained over a period of several hundred milliseconds, especially at parietal sites. Because the auditory stimulus had not yet been presented, these differences were instead elicited by the pictures themselves, and may thus reflect differences in the familiarity with known items relative to unknown items. For example, this pattern could be interpreted as an N400-like effect of semantic integration, such that increased difficulty with integrating the unknown pictures into semantic memory elicited the observed patterns. However, if this were the case, this would predict an increased negativity to unknown pictures over parietal sites, whereas we currently see an increased positivity in this region. Although this effect and its interpretation need further replication, the finding of familiarity differences to the pictures themselves would be a potentially

Table 4 Correlations between implicit measure dependent variables and behavioral measures

	PPVT		Forced-Choice RT		Congruity RT	
	Known	Unknown	Known	Unknown	Known	Unknown
EM						
Total number of fixations	–.13	–.27	$.41^{§}$	.29		
Mean fixation duration	$–.49^{*}$	–.29	$.57^{*}$	.09		
First fixation duration	$–.44^{*}$	–.13	$.56^{*}$	.12		
First dwell time on stimulus	–.34	.09	$.91^{**}$	$.43^{§}$		
Latency to first fixation	.15	–.01	$.42^{§}$	.11		
Latency to first refixation	–.26	$–.36^{§}$	–.17	.07		
Proportion of fixation duration on the stimulus	$–.35^{§}$	.02	$.49^{*}$	.14		
Proportion of dwell time on stimulus	$–.42^{*}$	.17	$.69^{**}$	.06		
First (%)	.02	–.04	–.23	–.06		
Last (%)	–.14	$.54^{**}$	.37	–.30		
PD						
Peak dilation	.02	.11	–.29	–.07		
Mean change in pupil size	–.14	.28	–.26	$–.72^{**}$		
Maximum percent change in pupil size	–.04	.10	–.14	–.01		
ERP						
N400 effect	$–.37^{§}$	–.33			.29	.11

Behavioral RTs are correlated with dependent measures that were collected during the same paradigm only. The congruity RT and N400 effects were based on incongruent – congruent differences; see the text for further explanation. Statistically significant correlations are indicated with asterisks: $^{§}p < .10$, $^{*}p < .05$, $^{**}p < .01$

Table 5 Correlations between all eye movement monitoring (EM), pupillometry, and event-related potential (ERP) measures, for known trials only

Known trials		Total number of fixations	Mean fixation duration	First fixation duration	First dwell time	Latency to first fixation	Latency to first refixation	Proportion of fixation duration on the stimulus	Proportion of dwell time on stimulus	First (%)	Last (%)	Peak dilation	Mean change in pupil size	Maximum percent change in pupil size	N400 effect
EM measures	Total number of fixations	1													
	Mean fixation duration	.11	1												
	First fixation duration	–.03	.96**	1											
	First dwell time	.37§	.74**	.66**	1										
	Latency to first fixation	–.46*	.13	.27	.22	1									
	Latency to first refixation	–.15	.18	.27	–.07	.18	1								
	Proportion of fixation duration on the stimulus	.12	.59**	.49*	.60**	.04	–.05	1							
	Proportion of dwell time on stimulus	.61**	.67**	.49	.85**	–.29	–.20	.62**	1						
	First (%)	–.69**	–.12	–.01	–.20	.53**	.30	.06	–.52*	1					
	Last (%)	.44*	.34	.08	.45*	–.34	–.37§	.46*	.71**	–.40§	1				
Pupil measures	Peak dilation	–.51*	–.04	.06	–.35	.21	.07	–.09	–.41	.28	–.07	1			
	Mean change in pupil size	–.18	–.13	–.11	–.21	.02	.35	–.14	–.23	.49*	–.18	.17	1		
	Maximum percent change in pupil size	–.65**	.15	.26	–.12	.33	.18	.19	–.28	.56**	–.10	.81**	.25	1	
ERP measure	N400 effect	.34	.43*	.45*	.39§	.21	.31	.08	.30	–.20	–.01	–.15	–.11	–.22	1

Statistically significant correlations are indicated with asterisks: § $p < .10$, * $p < .05$, ** $p < .01$.

Table 6 Correlations between all eye movement monitoring (EM), pupillometry, and event-related potential (ERP) measures, for unknown trials only

Unknown trials		Total number of fixations	Mean fixation duration	First fixation duration	First dwell time	Latency to first fixation	Latency to first refixation	Proportion of fixation duration on the stimulus	Proportion of dwell time on stimulus	First (%)	Last (%)	Peak dilation	Mean change in pupil size	Maximum percent change in pupil size	N400 effect
EM measures	Total number of fixations	1													
	Mean fixation duration	.09	1												
	First fixation duration	–.03	$.88^{**}$	1											
	First dwell time	$.35^{\S}$	.29	$.39^{\S}$	1										
	Latency to first fixation	–.34	–.10	.18	–.03	1									
	Latency to first refixation	–.03	$.54^{**}$	$.55^{**}$	–.09	.33	1								
	Proportion of fixation duration on the stimulus	.25	.15	.14	$.47^{*}$	–.15	–.15	1							
	Proportion of dwell time on stimulus	$.51^{*}$	.24	.09	$.67^{**}$	$–.64^{**}$	–.23	$.47^{*}$	1						
	First (%)	–.14	–.09	–.13	.05	.03	.07	–.01	.08	1					
	Last (%)	–.13	.06	–.05	.01	–.34	–.01	.01	$.46^{*}$	–.04	1				
Pupillometry measures	Peak dilation	$–.64^{**}$	.19	.28	–.21	.25	.13	–.13	–.30	.10	.26	1			
	Mean change in pupil size	–.30	–.15	–.11	.00	–.01	–.20	–.02	.08	.20	–.07	–.05	1		
	Maximum percent change in pupil size	$–.70^{**}$	.11	.24	–.08	$.41^{\S}$	.22	.05	–.31	.25	.10	$.79^{**}$	.13	1	
ERP measures	N400 effect	.27	.21	.25	.34	.08	–.08	.17	.08	–.17	–.35	–.25	–.13	–.35	1

Statistically significant correlations are indicated with asterisks: $^{\S}p < .10$, $^{*}p < .05$, $^{**}p < .01$

important addition to the body of research on the use of ERPs to study receptive vocabulary knowledge.

Importantly, the results from our two additional implicit assessment techniques showed similar patterns of reliable differentiation between the processing of known and unknown words. During a forced-choice task, we simultaneously collected EM and pupillometry data. In the EM data, participants' eye movements were generally faster to and more consistently focused on the named picture for known than for unknown words. The fixation duration measures (including mean fixation duration, first fixation duration, and first dwell duration) were all longer for known than for unknown words, demonstrating that participants spent more time looking at the named picture when it was associated with a known word than when it was associated with an unknown word. Participants were faster to move their eyes to the named picture for the first time (latency to first fixation) and to move their eyes back to the named picture after having left it (latency to first refixation) for known than for unknown words. Finally, proportional measures showed a similar finding: Proportions of fixations on the stimulus and proportions of dwell time on the stimulus showed that participants looked more at the named picture, and for longer amounts of time, in the known than in the unknown condition. Thus, reliable differences were observed across all dependent measures of eye movements, suggesting that the processing of known words was different than that of unknown words, in ways that suggest that identifying the named picture was easier for participants in the known condition. Similar findings were observed in the pupillometry data, in which all three of our dependent measures (peak dilation, average change in pupil size, and percent change in pupillary dilation) showed larger changes in the unknown than in the known condition. Because pupillary dilation has been shown to increase with task difficulty across a range of tasks in previous research, these results again support the conclusion that processing in the forced-choice task paradigm was more difficult for words that were expected to be unknown by our participants.

Our ERP results, then, replicated previous findings of reliable differences in processing measures of words depending on their receptive vocabulary status (known or unknown). Our EM and pupillometry results demonstrated that such differences can also be identified using other implicit assessment techniques that, to our knowledge, have not previously been used explicitly to study receptive vocabulary knowledge in any participant group. Thus, the EM and pupillometry results provide important confirmation of the ERP differences that were observed, with the similar advantage of not having to rely on explicit behavioral responses from participants. Crucially, although we collected behavioral measures in the present study, all three of these implicit measures yield reliable differences between known and unknown words that do not rely on these behavioral responses. These implicit techniques are thus invaluable for studying receptive vocabulary knowledge in populations who do not, or cannot, make overt responses.

We also looked at the correlations between our implicit measures and behavioral measures, as well as the correlations between the various implicit measures themselves. We observed significant relationships between a standardized measure of vocabulary (the PPVT) and several of our EM measures; a marginally significant correlation was also observed between PPVT score and N400 effect size. We also observed significant correlations between RTs on the forced-choice task and a number of EM measures. We did not observe significant relationships between the N400 effect size and the RT difference measure on the congruity task; this may have been due in part to a lack of statistical power, due to the inherently noisy ERP data.

We observed a number of significant correlations between measures from the different assessment techniques, as can be seen, for example, in the positive correlations between N400 effect size and some of the EM duration measures for known words. However, in a number of cases significant correlations were not observed across the measures. In some cases, this may have been due to a lack of statistical power, due to the inherent noisiness of the ERP data; for example, negative correlations between N400 effect size and the pupillary measures (such that larger N400 differences were observed when pupillary dilation changes were smaller, and vice versa) were observed for both the known and unknown words, but these correlations did not reach significance. Other cases, however, may be due to the fact that the implicit measures reflect underlying processes that, though all engaged in the service of word and picture recognition, potentially vary quite widely in their exact cognitive function. These differences may be exacerbated by the differences in task requirements for the forced-choice and congruity paradigms. What is important for the present purposes is that even when the implicit measures did not correlate with each other (perhaps because they were tapping into complementary but nonoverlapping cognitive processes), all three were still capable of differentiating groups of known words from groups of unknown words.

These results suggest that eye movements and pupillometry might provide techniques alternative to ERPs for the assessment of receptive vocabulary knowledge. The availability of alternatives to ERPs for such testing might be appealing for several reasons, as we proposed in the introduction. First, EM and pupillometry might be available to researchers or clinicians for whom the ERP methodology is not (for practical, financial, or other reasons). Being able to use EM or pupillometry might thus make the implicit assessment of receptive vocabulary more accessible to a wider group of individuals for whom such knowledge might prove useful (e.g., for research purposes or for the purposes of developing rehabilitative therapy). Second, EM and pupillometry

recording might be accomplished better with some participant groups than are ERPs. Many eyetracking systems are capable of collecting EM and pupillary data in a noninvasive manner, without the need for equipment that touches the participant in any way. This is not the case for ERP research, which requires the placement of electrodes on participants' scalps. Even under the most ideal circumstances—for example, with modern recording systems that try to minimize the time required for electrode application—the process of applying the electrodes to the scalp, and the necessity of keeping them there during data acquisition, can prove difficult with some participant groups (such as small children and infants), and perhaps can be highly stressful to others (such as individuals with autism, who have demonstrated sensitivities to and varying tolerances for objects placed on their person). Similarly, the desire to minimize EEG artifacts during data acquisition is more easily accomplished with some participant groups than with others; although eyetracking systems have their own constraints in this regard, some participant groups may be better able to comply with the eyetracking restrictions than with those imposed by ERPs.

These results suggest that in the future, the use of eyetracking measures (EM and pupil dilation) might help to overcome one of the greatest obstacles to using ERPs to better understand receptive vocabulary: the need to average the dependent measure across multiple items in ERP experiments. This makes it difficult to examine the event-related signals to individual items, such as vocabulary words. Ideally, though, this might be *exactly* what we would most like to do: to make a determination, on the basis of the response to an individual item, whether that specific item is likely to be known or unknown to the individual participant. Such information would be invaluable to clinicians and teachers in developing and personalizing instruction. The use of EM and pupillometry might offer just such a possibility, since these measures do not depend on signal averaging across trials of like types. In fact, the multiple dependent measures that can be recorded using either EM or pupillometry might have the additional advantage of allowing the trained evaluator to look at patterns across dependent measures for a single vocabulary word to make a determination about whether that word is or is not known to an individual participant. We have explored these possibilities by using ERP, EM, and pupillometry measures to model subjective knowledge ratings using mixed-effects logistic regression (Coderre, Gordon, & Ledoux, under review).

Finally, the use of any one of these implicit measurement techniques (EM, pupillometry, or ERPs) may prove especially advantageous to the study of receptive vocabulary knowledge in participant groups from whom reliable overt behavioral responses are difficult to collect. These very frequently may be the very participant groups in which such knowledge could be most beneficial in terms of further learning. In ongoing work, we have been using these same paradigms to collect implicit measures of receptive vocabulary knowledge in typically developing children and in high- and low-functioning individuals with autism. This last group, in particular, has been especially difficult to study using more traditional measurement techniques. Given the pervasive language and communication deficits observed for low-functioning individuals with autism, having measures of what words may or may not be understood by individual participants could prove useful in terms of further instruction and in terms of improving caregiver communication.

Author note This research was supported by grants from the Nancy Lurie Marks Foundation, the Department of Defense Autism Research Program (Grant No. AR093137), the Therapeutic Cognitive Neuroscience Fund, and the Benjamin and Adith Miller Family Endowment on Aging, Alzheimer's, and Autism Research. E.B. is now at the University of Rochester; I.G. is now at the University of Wisconsin.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439. doi:10.1006/jmla.1997.2558

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276–292. doi:10.1037/0033-2909.91.2.276

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). New York, NY: Cambridge University Press.

Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5,* 34–44.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. doi:10.3758/BRM.41.4.977

Byrne, J. M., Connolly, J. F., MacLean, S. E., Dooley, J. M., Gordon, K. E., & Beattie, T. L. (1999). Brain activity and language assessment using event-related potentials: Development of a clinical protocol. *Developmental Medicine and Child Neurology, 41,* 740–747.

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory, 92,* 225–234.

Connolly, J. F., & D'Arcy, R. C. N. (2000). Innovations in neuropsychological assessment using event-related brain potentials. *International Journal of Psychophysiology, 37,* 31–47.

Connolly, J. F., D'Arcy, R. C. N., Newman, R. L., & Kemps, R. (2000). The application of cognitive event-related brain potentials (ERPs) in language-impaired individuals: Review and case studies. *International Journal of Psychophysiology, 38,* 55–70.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology, 6,* 84–107.

D'Arcy, R. C. N., Marchand, Y., Eskes, G. A., Harrison, E. R., Phillips, S. J., Major, A., & Connolly, J. F. (2003). Electrophysiological assessment of language function following stroke. *Clinical Neurophysiology, 114,* 662–672.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access:

Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507–534.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken word recognition. *Psychonomic Bulletin & Review, 12,* 453–459. doi:10.3758/BF03193787

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30,* 412–431. doi:10.1016/0001-6918(69)90065-1

Dunn, L. M., & Dunn, L. M. (2007). *Peabody picture vocabulary tests* (4th ed.). Circle Pines, MN: American Guidance Service.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409–436.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146,* 668–670.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42,* 98–116.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72,* 1003–1015.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience, 16,* 1465–1477. doi:10.1162/0898929042304705

Friedrich, M., & Friederici, A. D. (2005a). Lexical priming and semantic integration reflected in the ERP of 14-month-olds. *NeuroReport, 16,* 653–656.

Friedrich, M., & Friederici, A. D. (2005b). Phonotactic knowledge and lexical–semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience, 17,* 1785–1802. doi:10.1162/089892905774589172

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *NeuroReport, 19,* 1757–1762.

Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language, 114,* 66–71.

Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological Bulletin, 77,* 340–355. doi:10.1037/h0032456

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary, and decoding. *Brain and Language, 117,* 88–99.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American, 212,* 46–54.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143,* 1190–1192.

Hess, E. H., Seltzer, A. L., & Shlien, J. M. (1965). Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology, 70,* 165–168.

Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers, 25,* 16–26. doi:10.3758/BF03204445

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30,* 47–61.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96,* B23–B32. doi:10.1016/j.cognition.2004.10.003

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology, 64,* 122–145.

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia, 50,* 3702–3712.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8,* 441–480. doi:10.1016/0010-0285(76)90015-3

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154,* 1583–1585.

Kahneman, D., Beatty, J., & Pollack, I. (1967). Perceptual deficit during a mental task. *Science, 157,* 218–219.

Karatekin, C., Marcus, D. J., & Couperus, J. W. (2007). Regulation of cognitive resources during sustained attention and working memory in 10-year-olds and adults. *Psychophysiology, 44,* 128–144.

Key, A. P. F., Molfese, D. L., & Ratajczak, E. D. (2006). ERP indicators of learning in adults. *Developmental Neuropsychology, 29,* 379–395.

Kuipers, J.-R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5,* 61. doi:10.3389/fnhum.2011.00061

Kuipers, J.-R., & Thierry, G. (2013). ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex, 49,* 2853–2860. doi:10.1016/j.cortex.2013.01.012

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463–470.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205. doi:10.1126/science.7350657

Loewenfeld, I. E. (1993). *The pupil: Anatomy, physiology, and clinical applications* (Vol. 1). Ames, IA: Iowa State University Press.

Marchand, Y., D'Arcy, R. C. N., & Connolly, J. F. (2002). Linking neurophysiological and neuropsychological measures for aphasia assessment. *Clinical Neurophysiology, 113,* 1715–1722.

Myung, J.-Y., Blumstein, S. E., Yee, E., Sedivy, J. C., Thompson-Schill, S. L., & Buxbaum, L. J. (2010). Impaired access to manipulation features in apraxia: Evidence from eyetracking and semantic judgment tasks. *Brain and Language, 112,* 101–112. doi:10.1016/j.bandl.2009.12.003

Myung, J.-Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition, 98,* 223–243. doi:10.1016/j.cognition.2004.11.010

Odekar, A., Hallowell, B., Kruse, H., Moates, D., & Lee, C.-Y. (2009). Validity of eye movement methods and indices for capturing semantic (associative) priming effects. *Journal of Speech, Language, and Hearing Research, 52,* 31–48.

Ojima, S., Nakamura, N., Matsuba-Kurita, H., Hoshino, T., & Hagiwara, H. (2010). Neural correlates of foreign-language learning in childhood: A 3-year longitudinal ERP study. *Journal of Cognitive Neuroscience, 23,* 183–199.

Posner, M. I. (2005). Timing the brain: Mental chronometry as a tool in neuroscience. *PLoS Biology, 3*(e51), 204–206. doi:10.1371/journal.pbio.0030051

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects in 24-month-olds. *Journal of Memory and Language, 46,* 39–56.

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 443–469). Amsterdam, The Netherlands: Elsevier.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557–580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. doi:10.1126/science.7777863

Torkildsen, J. V. K., Friis Hansen, H., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language, 1081,* 73–88. doi:10.1016/j.bandl.2008.09.005

Torkildsen, J. V. K., Svangstu, J. M., Friis Hansen, H., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience, 20,* 1266–1282. doi:10.1162/jocn.2008.20087

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89–134.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition, 19,* 95–112.

Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: Activation of shape and function features during object identification. *Journal of Experimental Psychology: General, 140,* 348–363. doi:10.1037/a0022840

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1–14. doi:10.1037/0278-7393.32.1.1

Appendix 2

Coderre, E., Chernenok, M., O'Grady, J., Bosley, L., Gordon, B., & Ledoux, K. (Under review.) Implicit measures of receptive vocabulary knowledge in low-functioning individuals with autism. *Journal of Speech, Language, & Hearing Research*.

Journal of Speech, Language, and Hearing Research

Implicit Measures of Receptive Vocabulary Knowledge in Low-Functioning Individuals With Autism

Journal:	*Journal of Speech, Language, and Hearing Research*
Manuscript ID	Draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Coderre, Emily; Johns Hopkins University School of Medicine, Neurology Chernenok, Mariya; Johns Hopkins University School of Medicine, Neurology O'Grady, Jessica; Johns Hopkins University School of Medicine, Neurology Bosley, Laura; Johns Hopkins University School of Medicine, Neurology Gordon, Barry; Johns Hopkins University School of Medicine, Neurology; Johns Hopkins University, Cognitive Science Ledoux, Kerry; Johns Hopkins University School of Medicine, Neurology
Keywords:	Autism, Cognition, Electrophysiology, Language

Implicit Measures of Receptive Vocabulary Knowledge in Low-Functioning Individuals With Autism

Emily L. Coderre[a], Mariya Chernenok[a], Jessica O'Grady[a], Laura Bosley[a], Barry Gordon[a,b], & Kerry Ledoux[a]

[a] Division of Cognitive Neurology/Neuropsychology, Department of Neurology

The Johns Hopkins University School of Medicine

1629 Thames St., Suite 350, Baltimore, MD, 21231, USA

[b] Department of Cognitive Science

The Johns Hopkins University

237 Krieger Hall, 3400 North Charles St., Baltimore, MD, 21218, USA

Address for correspondence (EC):

Emily Coderre

Johns Hopkins University School of Medicine

Cognitive Neurology/Neuropsychology-Department of Neurology

1629 Thames Street, Suite 350

Baltimore, MD 21231

Phone: +1 (443) 287-1716

Email: ecoderr1@jhmi.edu

Abstract

Purpose: Implicit measures of cognition are essential for assessing knowledge in low-functioning individuals with autism (LFAs), because such individuals are often unable to make reliable overt behavioral responses. Here we test whether three implicit measures – eye movement monitoring (EM), pupillary dilation (PD), and event-related potentials (ERPs) – can reliably estimate vocabulary knowledge in LFAs.

Methods: Five LFA adults were tested in a repeated-measures design with two tasks. High-frequency 'known' words (e.g. *bus*, *airplane*) and low-frequency "unknown" words (e.g. *ackee*, *cherimoya*) were presented in a visual-world task (during which EM and PD data were collected) and a picture-word congruity task (during which ERP data were collected).

Results: Using a case study approach with single-subject analyses, we demonstrate that these implicit measures can provide estimates of receptive vocabulary knowledge in the majority of these LFA participants. However, participants differed with respect to which measures were the most sensitive and which variables best predicted vocabulary knowledge.

Conclusions: These implicit measures may be useful to assess language abilities in LFAs, but their use should be tailored to each individual. This work holds important implications for the development of individualized implicit assessments of receptive vocabulary knowledge in populations unable to provide explicit behavioral responses.

Keywords: low-functioning autism; vocabulary; eye-tracking; pupillometry; ERP

Implicit Measures of Receptive Vocabulary Knowledge in Low-Functioning Individuals With Autism

Autism spectrum disorder (ASD) is a pervasive developmental disorder affecting one in 68 children (CDC, 2014). Language impairment is a hallmark characteristic: approximately 25% of individuals with ASD have little-to-no functional speech and are “non-verbal” (Turner, Stone, Pozdol, & Coonrod, 2006). Even in those with functional speech, production deficits are pervasive. While limitations in functional speech do not preclude functional comprehension, assessing comprehension in non-verbal individuals can be difficult since these individuals often exhibit low reliability in behavioral responses. Because of such difficulties with testing, lower-functioning individuals with autism are extremely under-represented in studies of cognition, making our knowledge of autism consequently incomplete. The demonstration that low-functioning individuals can comprehend language in the absence of functional speech would have broad clinical applications and could inform new approaches to remediation for individuals who struggle with traditional production-oriented techniques.

While obtaining overt reports of language abilities may be difficult in individuals with severe autism, implicit measures of language comprehension, which can be collected and interpreted in the absence of behavioral responses, may provide alternative assessments. The current work represents an exploratory, proof-of-concept study investigating eye movement (EM) monitoring, pupillary dilation (PD), and event-related potentials (ERPs) in assessing receptive vocabulary knowledge in five low-functioning individuals with autism (LFAs). Given challenges to testing and the inevitable heterogeneity of severe autism (see ‘Specific considerations for LFAs’), we

adopt a case study[1] methodology with single-subject analyses to demonstrate the utility of these implicit measures in individualized assessments and interventions.

Implicit measures of language

EMs, PD, and ERPs have been established as valid implicit measures of receptive language in typically-developing (TD) adults. The so-called "visual world paradigm", in which a visual display of pictures is followed by a spoken word or phrase, has become a canonical technique to assess online spoken language comprehension (Tanenhaus, Magnuson, Dahan, & Chambers, 2000; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Participants' eyes typically move toward a named picture as soon as it can be identified and disambiguated from other pictures. Similarly, when a written word is presented before the picture display (e.g. *marriage*), EMs are faster to a semantically-related picture (e.g. ring) compared to an unrelated picture (e.g. pencil; Odekar, Hallowell, Kruse, Moates, & Lee, 2009). Importantly, these EM patterns occur in the absence of a behavioral task (Odekar et al., 2009), indicating their utility as implicit measures of language comprehension.

PD (in keeping with the terminology in the pupillometry literature, 'dilation' is referred to here as an increase in pupil diameter) in response to a stimulus typically increases with cognitive load (Beatty & Lucero-Wagoner, 2000; Granholm, Asarnow, Sarkin, & Dykes, 1996). PD is thus taken to reflect resource recruitment and has been used to assess processing demands in numerous cognitive tasks (Beatty & Lucero-Wagoner, 2000). In language comprehension studies, unrelated pairs of pictures and spoken words (e.g. duck-"bed") elicit greater PD compared to matched pairs (e.g. duck-"duck"), indicating greater resource recruitment in unrelated conditions (Kuipers & Thierry, 2011, 2013). Such effects occur in the absence of a

behavioral task (Kuipers & Thierry, 2011), demonstrating the utility of PD as an implicit measure of language comprehension.

ERPs are derived by time-locking changes in the electroencephalogram (EEG) to a stimulus onset. Specific ERP components are associated with various aspects of language (Kutas, van Petten, & Kluender, 2006; Sereno & Rayner, 2003). For current purposes, the N400 component is taken to reflect semantic processing and integration (Kutas & Hillyard, 1980; Lau, Phillips, & Poeppel, 2008; see Kutas & Federmeier, 2011 for a broader discussion). N400 amplitude is reduced when a stimulus is easily integrated with its preceding context (e.g. semantically-related or congruent stimuli). This amplitude reduction, compared to conditions with more difficult semantic integration (e.g. semantically-unrelated or incongruent stimuli), is termed the "N400 effect" and is thought to index semantic integration. The N400 effect occurs in the absence of behavioral responses (Kuipers & Thierry, 2011), demonstrating its utility as an implicit measure of language comprehension. Importantly, however, the N400 effect is only elicited when the target concept is within an individual's vocabulary range (Byrne, Dywan, & Connolly, 1995; Connolly & D'Arcy, 2000). No N400 effect is observed for words unknown to the participant because prior knowledge cannot ease integration in these cases.

We have previously demonstrated the concurrent use of EMs, PD, and ERPs as implicit measures of receptive vocabulary knowledge in TD adults using high-frequency 'known' words (e.g. *airplane*) and low-frequency 'unknown' words (e.g. *cherimoya;* Ledoux et al., 2016). In a visual world paradigm, during which EM and PD data were collected, four pictures were followed by a spoken word matching one of the pictures. EM data indicated that participants could more quickly identify the target picture for 'known' than 'unknown' vocabulary: 'known' words had fewer fixations over the course of the trial, faster EMs to and longer fixations on the

target, and more trials on which the target was the last to be fixated on. Pupillometry results showed greater PD for ‘unknown’ than ‘known’ words, suggesting greater resource recruitment. In a separate session, ERP data were collected during a picture-word congruency paradigm in which a picture was followed by a spoken word that matched (congruent) or did not match (incongruent) the picture. An N400 effect (reduced N400 amplitude for congruent versus incongruent pairs) occurred for ‘known’ but not ‘unknown’ words, since participants could not evaluate the congruency of ‘unknown’ concepts. Overall, all three measures showed effects consistent with prior work using EMs, PD, and ERPs as implicit measures of language, demonstrating that they can be used in conjunction to assess receptive vocabulary in TD adults. Critically, although participants made behavioral responses throughout the tasks, all three implicit measures reliably distinguished between ‘known’ and ‘unknown’ words without relying on behavioral responses. The implicit nature of these measures makes them extremely valuable in studying cognition in populations unable to provide overt responses.

Implicit measures in autism

Prior research using implicit measures in individuals with ASD has documented notable differences between ASD and TD groups. Individuals with ASD show abnormal EM patterns during visual tasks (Brenner, Turner, & Müller, 2007; Goldberg et al., 2002; Mottron et al., 2007; Schmitt, Cook, Sweeney, & Mosconi, 2014) and atypical viewing patterns in visual world paradigms, such as lower proportions of looking time at the target picture (Bavin et al., 2014; Brock, Norbury, Einav, & Nation, 2008). Pupillometry studies have documented abnormalities such as larger baseline pupil size (Anderson & Colombo, 2009) and smaller change in pupil size in response to social stimuli (Martineau et al., 2011). ERP studies have reported reduced or

absent N400 effects in response to linguistic stimuli in individuals with ASD compared to TD individuals (Dunn et al. 1999; McCleery et al. 2010; Pijnacker et al. 2010).

However, most prior studies tested high-functioning individuals. Of those studies we reviewed, only Bavin et al. (2014) tested LFAs: In a visual world paradigm in children with ASD, including "severe autism," greater symptom severity was associated with lower proportions of looking time at target pictures. To our knowledge, the current study is the first to use these three implicit measures (EMs, PD, and ERPs) concurrently to assess receptive vocabulary knowledge in adolescent and adult LFAs.

Because of difficulties with testing and greater individual variability (see 'Specific considerations for LFAs'), using multiple measures is especially beneficial in lower-functioning populations (e.g. Plesa Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2015). Ledoux et al. (2016) demonstrated that EMs, PD, and ERPs could all reliably estimate receptive vocabulary, meaning that if one methodology is unavailable (e.g. a participant will not tolerate the EEG net or the presence of glasses makes eye-tracking difficult), the other(s) may provide an alternative. Similarly, we use multiple EM and PD variables since some may be better indices of receptive vocabulary than others in certain low-functioning individuals.

Given the documented atypical patterns in implicit measures in ASD, directly comparing LFAs to TD or higher-functioning groups could be problematic. Here we use assessment paradigms that have been widely validated with TD adults (including in Ledoux et al. (2016) using the same stimuli and methods). Critically, however, all current measures are within-subjects comparisons of 'known' and 'unknown' words. While atypical patterns of implicit measures may occur in ASD populations, these measures might still distinguish between 'known' and 'unknown' vocabulary *within* individuals. For instance, even if individuals with

ASD show reduced N400 effects compared to TD adults, an N400 effect may still be observed within *one* individual with ASD for 'known' but not 'unknown' words. The potential for these implicit measures to distinguish 'known' and 'unknown' vocabulary within-subjects would be highly informative regarding the utility of these measures to assess receptive vocabulary in LFAs.

Based on previous demonstrations that EMs, PD, and ERPs can differentiate 'known' from 'unknown' words in TD adults (Ledoux et al., 2016), the current exploratory, proof-of-concept study assessed whether these measures can also serve as reliable within-subject implicit measures[2] of receptive vocabulary knowledge in LFAs. To the extent that these measures would show similar patterns in TD adults and LFAs, we would predict similar results in the two groups: faster and more accurate EMs to 'known' words; greater PD to 'unknown' words; and an N400 effect for 'known' but not 'unknown' words.

Specific considerations for LFAs

Cognitive testing with LFAs is challenging due to idiosyncrasies of the autism disorder such as sensory abnormalities or difficulties understanding or following directions. For example, participants may be unable to use a response box or mouse, make responses haphazardly, or display no motivation to complete the task (e.g. Kylliäinen, Jones, Gomot, Warreyn, & Falck-Ytter, 2014). Such difficulties can lead to high rates of data loss, participant attrition, and/or increased variability in the data. Furthermore, some research suggests that the EEG activity of ASD participants is inherently noisier than in TD individuals (e.g. Pérez Velázquez & Galán, 2013; although see Davis & Plaisted-Grant, 2015), which may require collecting more EEG data to improve signal-to-noise ratios or making modifications to data acquisition or cleaning (e.g. Kylliäinen et al., 2014). These challenges with data acquisition and quality likely contribute to

the shortage of research performed with LFAs. Autism is also an extremely heterogeneous disorder with significant variation among individuals in terms of cognitive abilities, expressive and receptive language, and symptom severity – making it difficult to categorize individuals (see Participants section). While group analyses may be informative, single-subject examination is crucial, especially when testing low-functioning individuals.

Given these considerations for LFAs, and as the ultimate aim of this work is to determine vocabulary knowledge on an individual basis, we adopt a case study approach with single-subject statistical methods to assess the utility of EMs, PD, and ERPs in distinguishing 'known' and 'unknown' words in five LFAs. Single-subject analyses may elucidate which measures best predict language abilities in each participant, the strength of the effects, and the accuracy of each measure in estimating receptive vocabulary. Because implicit measures offer a promising method of accessing the latent constructs of language in low-functioning populations, this work is an important step in understanding cognition in LFAs, whose behavioral responses are often unreliable or unattainable.

Methods

Participants

Participants were five LFAs (mean age 32 years; all males; 4 Caucasian, 1 Asian) recruited from the Baltimore community. All had normal or corrected-to-normal vision and hearing, as assessed via caregiver or self-report. Experimental procedures were approved by the Johns Hopkins School of Medicine Institutional Review Board. For those participants who were unable to provide their own informed consent (DL and HD), we followed the Maryland law applicable to surrogate decision-making for health care, stating that a legal guardian may provide consent on behalf of the participant. For those participants who were able to legally provide their own

consent (WF, SE, and PB), we obtained written informed consent from the participants as well as from their group home benefits manager.

Although intellectual and verbal abilities are often used to determine functioning level, in the current study “low-functioning autism” was defined according to DSM-5 Level 3 (Severe Level of Autism), which marks severe deficits in social communication and restricted and repetitive behaviors requiring substantial support throughout the individual’s daily life. Criteria for identifying participants as LFAs were based on the severity of core features of autism as stated in DSM-5; the level of environmental support and supervision needed; and scores on the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000, 2012) and/or Autistic Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994). All participants exhibited restricted and repetitive behaviors and severe deficits in verbal and/or non-verbal social communication skills that significantly affected their level of daily functioning. Each participant required direct 24-hour support staff and/or parental supervision, with a focus on activities of daily living and functional communication. All were enrolled in adult or educational programs targeted to individuals with autism.

Neuropsychological assessments. All participants had a current diagnosis of autism, which was verified using the ADOS (First (ADOS-1) or Second Edition (ADOS-2), depending on the current version at the time of testing) and/or ADI-R. These assessments were administered by research team members who had completed the official ADOS clinical training. The Kaufman Brief Intelligence Test, Second Edition (K-BIT-2; Kaufman & Kaufman, 2004) and the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007) were administered to assess intelligence and receptive vocabulary, respectively. Although intelligence and language ability were included in obtaining an overall picture of each participant, they were noted as

possible associated features of autism and were not included in identifying these individuals as low-functioning. While all participants were classified as LFAs for the purposes of this study, they varied in their symptom severity, intelligence, and language abilities (Table 1 and Participant Descriptions section).

Table 1 shows test scores for each participant. For two participants there was no appropriate module of the ADOS, as the module that met criteria for expressive language skills was developmentally inappropriate for the participant's chronological age. The researchers performed "adapted" modules by interacting with these participants and identifying the specific behaviors measured by the ADOS. These adapted scores are noted in Table 1, but cannot be considered "official" ADOS scores.

Participant descriptions

DL. DL is an 18-year-old male who was diagnosed with autism at age 3. Developmental motor milestones were reached on schedule. At about 12 months, he started using five or six single words. At 18 months he regressed and stopped using verbal communication. DL is non-verbal and has no functional speech. He can communicate his needs by gesturing and using a topic board and single-symbol voice output device. He can comprehend language incorporated in his daily routine. He is hypersensitive to touch, displays stereotyped motor behaviors, and has impaired fine motor and visual motor skills. He can eat, use the toilet, and dress himself independently. Because DL is a non-verbal adult, an adapted ADOS-1 Module 1 was used. An ADI-R was also completed. Both assessments confirmed the diagnosis of autism. DL was unable to complete the PPVT or K-BIT due to an inability to understand test directions.

HD. HD is a 15-year-old male with a diagnosis of autism. He was diagnosed with a speech disorder when he was not talking at 15 months. Between ages 13 and 14, he began using

occasional single words, and at age 14 began using occasional simple phrases like “I want.” He is functionally non-verbal and produces little spontaneous speech, but can communicate with gestures and a visual communication system. He can understand simple commands and has a receptive vocabulary of approximately 50 words. HD displays stereotyped motor behaviors like hand flapping. An ADOS could not be completed at the time of initial assessment, and HD was thereafter unavailable for testing. An ADI-R was completed, which confirmed the diagnosis of autism. HD was able to complete some of the PPVT, but was unable to complete a K-BIT due to lack of compliance and an inability to understand test directions.

WF. WF is a 39-year-old male with diagnoses of autism and OCD. Records on WF’s early language development and initial diagnosis were not available at his current residential facility. Additional attempts to track previous records were unsuccessful. WF often uses stereotyped/idiosyncratic words and phrases that significantly impair his language and communication. He demonstrates a compulsion of rituals like a verbal routine he performs before responding to a social initiation. He experiences aggressive, hyperactive, and distractible behaviors, which have been treated with medication. He has stereotyped motor behaviors, including his walking patterns and hand movements. WF lives in a group home and is capable of all daily living activities. He assists with chores, participates in group home activities, works part time at a day program building jump ropes and packaging toys, and volunteers at the Red Cross packaging promotional materials. WF is an adult with functional speech, but because his speech is stilted and difficult to understand an adapted ADOS-2 Module 4 was used, which confirmed the diagnosis of autism. An ADI-R was not performed because WF’s legal guardians could not provide information about his infancy and early development (a major component of the ADI-R),

and his parents were unavailable for contact. His K-BIT and PPVT scores indicated verbal abilities in the range of intellectual dysfunction but unimpaired non-verbal abilities.

SE. SE is a 40-year-old male with a diagnosis of autism. Records on SE's early language development and initial diagnosis were not available at his current residential facility. Additional attempts to track previous records were unsuccessful. SE is outgoing and memorizes many things in regards to people, including names. His speech is fluent but includes many stereotypes and high amounts of scripting. He exhibits excessive interest in highly specific topics. SE lives in a group home and is capable of all daily living activities. He assists with chores and works part time at a boutique where he empties trash, at a school cleaning bathrooms, and at a landfill. SE is an adult with functional speech so an ADOS-2 Module 4 was used, which confirmed the diagnosis of autism. An ADI-R was not performed because SE's legal guardians could not provide information about his infancy and early development, and his parents were unavailable for contact. SE had difficulty maintaining attention and understanding task instructions for the K-BIT and PPVT; therefore his scores, which indicated intelligence and verbal abilities in the range of intellectual dysfunction, may not be an accurate reflection of his true abilities.

PB. PB is a 48-year-old male with a diagnosis of autism. Records on PB's early language development and initial diagnosis were not available at his current residential facility. Additional attempts to track previous records were unsuccessful. PB speaks quietly and often mumbles but has fluent speech. He has severe repetitive behaviors, which are compulsive and ritualistic in nature and interfere with his ability to function in society. Certain spoken sounds or phrases elicit repetitive or self-injurious behaviors. PB has become visibly agitated or aggressive when these compulsive behaviors are interrupted. PB lives in a group home and is capable of all daily living activities. He assists with chores, participates in group home activities, and works part time at a

landfill. He can manage small amounts of money independently. PB is an adult with functional speech so an ADOS-2 Module 4 was used, which confirmed the diagnosis of autism. An ADI-R was not performed because PB's legal guardians could not provide information about his infancy and early development, and his parents were unavailable for contact. His K-BIT and PPVT scores did not indicate intellectual impairment.[3]

Stimuli

Stimuli consisted of 160 auditory words and matching pictures (Figure 1 and Supplementary Material). Half were high-frequency words (average frequency per million in the Corpus of Contemporary American English (Davies, 2008)=56.5, *SD*=84.1) like *bus*. These were classified as 'known,' as we expected most of these words to be known by the participants. Half were extremely low-frequency words (average frequency per million=0.4, *SD*=0.7; although given their low frequency, many do not occur in language corpora), like *avocet*, which were classified as 'unknown' and expected to be unfamiliar to participants. 'Unknown' words had slightly more letters (*M*=6.8, *SD*=1.6) than 'known' words (*M*=5.1, *SD*=1.5). In addition to these objective classifications[4], each participant's parent or caregiver subjectively rated whether the individual knew each word receptively. These ratings estimated that all participants were familiar with most of the 'known' words and unfamiliar with all of the 'unknown' words in this stimulus set.

For picture stimuli, high-resolution color photos were selected from online sources to represent each word (Figure 1 and Supplementary Material). Pretesting with three TD adults confirmed these images represented the corresponding concepts (dictionary definitions were provided for 'unknown' words). All words were highly imageable, as determined through pretesting. Picture luminance was matched across 'known' and 'unknown' words. For auditory stimuli, high-quality auditory recordings were made for each word using Audacity 1.3 and edited

using Computerized Speech Lab Model 4150 (KayPENTAX). Auditory stimuli ranged from 500-1200 ms duration.

Task Procedure

The experiment consisted of a visual world task (EM and PD) and a picture-word congruity task (ERP), completed in separate sessions. Some participants underwent multiple sessions per task to ensure adequate amounts of usable data (see Table 2 and 'Number of Sessions' section).

Visual world task. The visual world paradigm was presented in E-Prime 2.0.8.74. In each trial, a central fixation cross was presented for 1000 ms. Four pictures were then presented, one centered in each quadrant, followed 20 milliseconds (ms) later by an auditory word. 'Known' words were always presented with 'known' distractors, and 'unknown' words with 'unknown' distractors, so participants could not eliminate foils in the 'unknown' condition based on familiarity. All four pictures remained on the screen for a maximum of 5000 ms after word presentation or until the participant selected a picture with a mouse click. These stimulus parameters are similar to previous studies using the visual world paradigm or obtaining PD measures (Kuipers & Thierry, 2011; Odekar et al., 2009). The experimental session consisted of 160 pseudorandomized trials (one per item) in 8 blocks of 20 trials each. Pictures were presented at 1.6-9.5^{o} of visual angle on a MicroTouch 3M 15" LCD monitor with 1024x768 resolution. EM and PD data were collected using an ASL Model 504 eye-tracking system. Pupil diameter was measured horizontally and recorded every 17 ms in pixels. The entire session lasted approximately 30 minutes, including approximately 15 minutes for equipment setup and calibration. To maintain attention, participants were asked to indicate, using the computer mouse, which picture matched the spoken word. All participants made behavioral responses; however, only WF, SE, and PB were able to understand the task instructions.

Picture-word congruity task. The picture-word congruency paradigm was presented in E-Prime. A centrally-presented picture was followed 700 ms later by a spoken word. Each word was presented twice: once in an incongruent (word and picture did not match) and once in a congruent condition (word and picture matched), yielding 320 trials total. Incongruent picture-word pairs were drawn from the same knowledge condition ('known' or 'unknown') and did not share an initial phoneme. The picture was presented for 1000 ms after the offset of the auditory stimulus. Pictures were presented at 2.4-9.5^{o} of visual angle on a Dell 17" LCD monitor with 1280x1024 resolution. ERPs were recorded at 250 Hz using a 256-channel Hydrocel Geodesic Sensor Net and NetStation 4.3. Impedences were kept under 50kΩ. Videos were recorded from the front and back to code for any "bad" trials during data preprocessing (see Data Preprocessing). The entire session lasted approximately 35 minutes, including approximately 15 minutes for net application and setup. The behavioral task for this paradigm required participants to withhold their response until a delayed fixation cross appeared (to minimize movement artifacts) and then indicate whether the word and picture matched using a button press. DL, HD, WF, and SE did not understand these instructions and did not make behavioral responses. PB understood the instructions but was unable to reliably wait for the response fixation cross, so the majority of his responses were not captured.

Number of sessions. Table 2 shows the number of trials collected and used in the final analyses. We required that approximately half of the total trials collected for each measure be usable. DL was unable to complete an entire eye-tracking session, with 120 trials collected. Due to excessive movement in the first EEG session, DL performed two additional shorter sessions approximately a month later. HD performed one session each of the eye-tracking and EEG tasks. WF performed one eye-tracking session; due to movement and noise artifacts in the first EEG

session, a second session was performed approximately 2 months later. SE required two eye-tracking and two EEG sessions due to excessive movement, talking, and difficulty with compliance in the first sessions; second sessions were performed approximately 2 months later. PB performed one eye-tracking session; due to excessive movement during the first EEG session, a second session was run approximately one month later.

Data Preprocessing

Eye movement data. EM data from the visual world task were analyzed using ASL Results (Applied Science Laboratories, 2009). Each visual display was divided into five regions of interest (ROIs), consisting of the four pictures and the central fixation. The 'target' is referred to here as the named picture on each trial. A fixation was operationalized as a time period during which eye gaze remained at one location. A stable gaze duration for 100 ms or more and a visual angle variation of $\leq 1^o$ determined a fixation onset. Three or more sequential fixations deviating from the onset location by $\geq 1^o$ of visual angle determined a fixation offset. Dwell time was operationalized as the time spent looking at the target, with or without fixation. If less than half of the trial was detected by the eye-tracker (i.e. the sum of all fixation durations was <50% of the total trial length), that trial was removed.

In Ledoux et al. (2016), all of the EM variables examined showed significant differences between 'known' and 'unknown' words in TD adults. In this study of LFAs, all EM variables were included, since some variables might be better indices of receptive knowledge than others. For each trial, the following variables were calculated. All duration and latency measures are in milliseconds.

Total number of fixations: total number of fixations in entire trial.

Mean fixation duration: average duration of fixations in target ROI.

First fixation duration: duration of first fixation in target ROI.

First dwell on stimulus: total time spent in target ROI, with or without fixation, during first entry.

Latency to first fixation: time elapsed before first fixation in target ROI.

Latency to first refixation: time elapsed before first refixation in target ROI (i.e., time to come back to target ROI after leaving target ROI).

Percentage of fixation duration on target: total fixation duration on target divided by total fixation duration for all pictures.

Percentage of dwell time on target: percentage of time spent in target ROI, with or without fixation (i.e., total dwell time on target/length of trial).

Percentage of trials first fixated: percentage of trials on which target was first picture fixated.

Percentage of trials last fixated: percentage of trials on which target was last picture fixated.

Because some participants had longer reaction times (RTs) for 'unknown' than 'known' trials (see Results), and because trials ended upon response (see Task Procedure), 'known' trials were sometimes shorter than 'unknown' trials. Differences in trial length would likely not impact latency measures (e.g. *latency to first fixation*); percentage measures, which divide by trial length, account for this difference automatically. *Number of fixations* is necessarily dependent on trial length and, as seen in the EM data, is often larger for 'unknown' than 'known' trials.

Pupillometry data. Pupillometry data from the visual world task were exported from ASL Results and analyzed in R (R Core Team, 2015). Pupil diameter was converted to millimeters and blinks were replaced by linear interpolation. For each trial, a 'baseline' pupil diameter (obtained by averaging over the 200 ms pre-stimulus time window) was subtracted from each measurement following stimulus presentation. Based on the pupillometry variables used in

Ledoux et al. (2016), three measures were calculated: *peak dilation*, *mean dilation*, and *maximum percent dilation*. Trials in which 20 or more consecutive data points (340 ms or more) were missing due to lack of fixations were removed.

ERP data. ERP data were preprocessed using EEGlab 10.2.2 (Delorme & Makeig, 2004) and Matlab 8.1 (MathWorks, Inc.). The data were bandpass filtered from 0.1-30 Hz and transformed to the average reference. Continuous data were segmented from 800 ms before to 1000 ms after the word (with the picture presented at -700 ms). Videos recorded during the EEG session were reviewed to identify and remove any "bad" trials containing movement, speaking, or inattention to the stimulus (e.g. not looking at the screen). Artifact correction was performed using independent component analysis (ICA; Delorme, Sejnowski, & Makeig, 2007; Jung et al., 2000). For participants with multiple sessions, the mean of each trial was removed before concatenating the sessions for ICA (Delorme & Makeig, 2004; Groppe, Makeig, & Kutas, 2009). Prior to ICA decomposition, the data were reduced to 64 dimensions. ICA components were reviewed individually and those contributing to sources of noise were removed from the data. Following ICA, a joint probability algorithm removed trials in which the amplitude at any channel or timepoint exceeded 3 standard deviations above or below the average amplitude for that channel. Finally, the cleaned data were visually reviewed, and any further bad trials (e.g. those containing artifacts not eliminated by the joint probability algorithm) were removed.

Statistical Analyses

Single-subject statistical analyses were performed in R using permutation tests. In the behavioral, EM, and PD data, all of a participant's individual trials were permuted to create a distribution of simulated test statistics. For each variable, 5,000 iterations were performed (which can estimate an alpha level of 0.01 to within 2%; Groppe, Urbach, & Kutas, 2011; Manly, 1997).

At each iteration, we permuted 'known' or 'unknown' labels between trials and ran a one-way (*trial type*: 'known'/'unknown') ANOVA. This repeated-measures approach accounts for the intercorrelation of the data, which are not independent nor paired across trials. The *F*-statistics from each iteration were used to create a null distribution from which the critical *F*-value corresponding to an alpha of 0.05 was calculated. We compared the observed *F*-value to the critical *F*-value to determine statistical significance. For observed *F*-values exceeding the critical *F*-values, *p*-values were derived for the observed effect. Bonferroni corrections for multiple comparisons were performed for the number of variables in each measure (10 in EM, 3 in PD). All reported *p*-values are Bonferroni-corrected unless otherwise specified.

For the ERP data, nine topographic regions were defined (clustered around F3/Fz/F4, C3/Cz/C4, and P3/Pz/P4; Figure 2). Data were collapsed over all electrodes within each cluster. Congruent vs. incongruent comparisons were performed separately for 'known' and 'unknown' trials. Based on previous literature, we would expect an N400 effect from approximately 300-500 ms after word onset. However, since no previous studies investigated N400 effects in LFAs, it is unclear whether latency differences would occur in this population. Rather than restrict analyses to pre-defined time windows, permutation tests were performed at every timepoint. To reduce the number of comparisons (Groppe et al., 2011), the data were downsampled to 125 Hz (one sample every 8 ms) and analyses were restricted to a time window from 200 ms after word onset (as congruency differences should not occur earlier than this) until the trial end. For each iteration, one-way (*congruency*: congruent/incongruent) ANOVAs were performed at each timepoint and electrode. Correction for multiple comparisons was performed using a cluster-based FWE correction at $p<0.05$ (full details in Groppe et al., 2011). Temporal clusters were defined as two or more consecutive timepoints showing effects at $p<0.05$. For each temporal

cluster, *F*-values were summed to obtain the cluster "mass". The largest cluster-level *F*-mass from each iteration was used to create a null distribution from which we derived the critical cluster *F*-mass corresponding to an alpha of 0.05. We then compared each observed cluster-level *F*-mass to the critical cluster *F*-mass to determine statistical significance.

Results

DL

In the behavioral data (Table 3), neither accuracy nor RTs on the visual world task differed significantly between 'known' and 'unknown' words (all p's > .82). Behavioral data were unavailable for the picture-word congruity task because DL could not understand task directions and did not provide behavioral responses.

No significant differences between 'known' and 'unknown' words occurred in the EM variables (all p's > .53; Figure 3a) or PD variables (all p's > .32 uncorrected; Figure 3b).

In the ERP data (Figure 3c), significant differences between congruent and incongruent conditions occurred in 'unknown' words at the Pz cluster from approximately 700-1000 ms. Congruity differences for 'unknown' words were unexpected; however, because DL showed congruity differences in 'unknown' words throughout the entire trial, we are inclined to attribute this finding to noise rather than a genuine N400 effect.

HD

In the behavioral data (Table 3), neither accuracy nor RTs on the visual world task differed significantly between 'known' and 'unknown' words (all p's > .20). Behavioral data for the picture-word congruity task were unavailable because HD could not understand task directions and did not provide behavioral responses.

No significant differences between 'known' and 'unknown' words occurred in the EM variables (all p's > .16; Figure 4a) or PD variables (all p's > .14; Figure 4b).

In the ERP data (Figure 4c) no significant differences between congruent and incongruent conditions occurred in either word type.

WF

In the behavioral data (Table 3), 'known' words showed significantly higher accuracy and faster RTs on the visual world task compared to 'unknown' words (all p's < .0001). Behavioral data for the picture-word congruity task were unavailable because WF could not understand task directions and did not provide behavioral responses.

In the EM variables (Figure 5a) 'known' words showed larger *mean fixation duration* ($F(1, 132) = 62.40$, $p < .01$); *first fixation duration* ($F(1, 127) = 8.52$, $p < .05$)[5]; *first dwell* ($F(1, 127) = 63.54$, $p < .01$); *percent fixation duration* ($F(1, 132) = 230.20$, $p < .01$); and *percent last fixated* ($F(1, 132) = 60.11$, $p < .01$) compared to 'unknown' words. 'Unknown' words showed a larger *number of fixations* ($F(1, 132) = 100.50$, $p < .01$) than 'known' words.

In the PD variables (Figure 5b), no significant differences between 'known' and 'unknown' words occurred (all p's > .47 uncorrected).

In the ERP data (Figure 5c), a significant N400 effect (incongruent more negative than congruent) occurred at the Pz cluster from approximately 200-400 ms. This effect occurred only for 'known' words; no congruency effects occurred for 'unknown' words.

SE

In the behavioral data (Table 3), 'known' words showed significantly higher accuracy and faster RTs on the visual world task compared to 'unknown' words (all p's < .05). Behavioral data

for the picture-word congruity task were unavailable because SE could not understand task directions and did not provide behavioral responses.

In the EM variables (Figure 6a), significant differences between ‘known’ and ‘unknown’ words occurred for *percent last fixated* ($F(1, 87) = 10.62$, $p < .05$), with ‘known’ words more often the last picture fixated compared to ‘unknown’ words.

In the PD variables (Figure 6b), no significant differences between ‘known’ and ‘unknown’ words occurred (all p’s > .14).

In the ERP data (Figure 6c), no significant differences between congruent and incongruent conditions occurred in either word type.

PB

In the behavioral data (Table 3), ‘known’ words showed significantly higher accuracy and faster RTs on the visual world task compared to ‘unknown’ words (all p’s < .0001). For the picture-word congruity task, responses were not recorded on the majority of trials (see Task Procedure section). Because not enough reliable data were available for analysis, PB’s behavioral data from this task were not analyzed.

In the EM variables (Figure 7a), ‘known’ words showed larger *percent fixation duration on stimulus* ($F(1, 112) = 75.89$, $p < .01$); *percent dwell* ($F(1, 112) = 64.47$, $p < .01$); and *percent last fixated* ($F(1, 112) = 72.42$, $p < .01$) compared to ‘unknown’ words. ‘Unknown’ words showed a larger *number of fixations* ($F(1, 112) = 78.49$, $p < .01$) than ‘known’ words.

In the PD variables (Figure 7b), ‘unknown’ words showed a trend toward larger *peak dilation* ($F(1, 120) = 4.50$, $p = .10$) and significantly larger *max percent dilation* ($F(1, 120) = 5.86$, $p < .05$) than ‘known’ words.

In the ERP data (Figure 7c), a significant N400 effect (incongruent more negative than congruent) occurred at the C3 cluster from approximately 400-550 ms. This effect occurred only for 'known' words; no congruency effects occurred for 'unknown' words.

Comparison of individual patterns

Table 4 summarizes each participant's results for each measure. To illustrate effect magnitudes for each variable and participant, within-subject 'unknown'-'known' differences for each variable were scaled to normalized *z*-scores (Figure 8a). Normalization within subjects enables comparison of effects on different scales and illustrates the strength of each effect in each participant. Topographic plots of incongruent-congruent differences illustrate ERP effects for 'known' and 'unknown' words (Figure 8b). Figure 8 demonstrates the variability within and between participants with regards to which measure(s) best distinguished between 'known' and 'unknown' words. For example, for WF, EM measures showed much larger effects than PD measures. Likewise, *mean fixation duration* was the largest effect for HD but showed no effect for DL. This variability also occurred in the EEG data: while WF and PB showed large N400 effects, the other participants showed negligible effects. Overall, Figure 8 illustrates that the specific measures that best elicit differences between 'known' and 'unknown' vocabulary may differ between individuals.

Discussion

Using a case study approach, this work investigated whether three implicit measures – EMs, PD, and ERPs – could provide within-subject assessments of receptive vocabulary knowledge in five LFAs. Based on previous results in TD adults (Ledoux et al., 2016), we predicted faster EMs and longer fixation durations, smaller PD, and larger N400 effects for 'known' words compared

to 'unknown' words. The results revealed notable differences among LFA participants in terms of which variables, if any, distinguished between 'known' and 'unknown' words.

Eye-movement monitoring

In Ledoux et al. (2016), all EM measures showed differences between 'known' and 'unknown' words. Here, only WF, SE, and PB showed significant effects on a subset of the EM variables. All three showed significant differences in *percent last fixated*. WF and PB showed differences in *number of fixations* and *percent fixation duration*. Only WF showed differences in *average fixation duration, first fixation duration*, and *first dwell*. These effects all replicated those found in TD adults (Ledoux et al., 2016). Importantly, the fact that some convergence occurred across participants in the measures eliciting significant differences may indicate that certain EM variables (specifically *percent last fixated*, *number of fixations,* and *percent fixation duration*) may be more sensitive in distinguishing 'known' and 'unknown' words. These measures may be the most valuable for future studies utilizing this paradigm to assess vocabulary knowledge in low-functioning populations.

In comparison, some variables (particularly latency measures) were less sensitive in distinguishing 'known' and 'unknown' words across participants. The non-significant effects in latency measures in LFAs could reflect baseline abnormalities in EM patterns. For example, Schmitt et al. (2014) observed slower, longer, and less-accurate saccades in individuals with ASD. Such baseline differences could have minimized 'known' and 'unknown' differences in the EM latency measures. We also observed no differences in *percentage of trials first fixated*, which could be explained by other idiosyncratic EM patterns in individuals with ASD such as strategic viewing patterns: LFA participants may be more likely to scan all pictures in the same

order on every trial (e.g. top-left, top-right, bottom-left, bottom-right) before returning to dwell on the target.

DL and HD showed no significant effects in any EM measures. However, trends in the expected direction were observed for *first fixation duration* and *percentage of trials last fixated* in DL and for *number of fixations*, *mean fixation duration*, *percent fixation duration*, *percent dwell*, and *percentage of trials last fixated* in HD. Interestingly, all five participants showed trends in the expected direction for *percentage of trials last fixated* (with statistically significant effects in WF, SE, and PB), which may suggest that this variable is the most informative measure for distinguishing 'known' and 'unknown' words, even if it does not show statistically significant differences in every participant.

Pupillary dilation

Only PB showed differences in the PD measures, specifically for *peak dilation* (although a statistical trend) and *max percent dilation.* These effects were larger for 'unknown' than 'known' words, replicating the pattern observed in TD adults (Ledoux et al., 2016). HD and WF showed non-significant trends in the expected direction for *mean dilation*. SE showed little difference between 'known' and 'unknown' words in any PD measures, and DL even showed a trend in the unexpected direction for *max percent dilation* (greater for 'known' than 'unknown'). Overall, these patterns suggest that the utility of PD measures for distinguishing 'known' and 'unknown' words differs between individuals.

Event-related potentials

WF and PB showed significant N400 effects only for 'known' words. These patterns replicate those observed in TD adults (Ledoux et al., 2016) and demonstrate that the N400 successfully distinguished between 'known' and 'unknown' vocabulary for these participants.

The N400 effect occurred at the Pz cluster from approximately 200-400 ms for WF and at the C3 cluster from approximately 400-550 ms for PB. The N400 typically occurs over centro-parietal scalp and anywhere from 200-600 ms in TD adults (Kutas & Federmeier, 2011). Thus, N400 topographies and latencies for WF and PB are consistent with previous literature.

HD and SE did not show N400 effects for either 'known' or 'unknown' words. DL showed a significant congruency effect only in 'unknown' words. However, because examination of DL's data indicated a sustained congruency effect over the entire trial for 'unknown' words, we are more inclined to interpret this effect as noise or drift than as an N400 effect. These findings suggest that ERPs may be better suited as implicit measures of receptive vocabulary in some participants than others.

Additional considerations

Overall, these results suggest that EMs, PD, and ERPs can provide implicit assessments of receptive vocabulary in LFAs, but that some measures are better suited for certain participants than for others (see also Plesa Skwerer et al., 2015). Only PB showed significant effects in all three measures; WF showed effects only in EM and ERP measures and SE only in EM measures. DL and HD showed no significant effects on any measures. Individual differences also occurred with regards to which variable(s) best distinguished 'known' and 'unknown' words. In the ERP data, variability occurred in the overall strength of the brain activity: Some participants had clear peaks in early perceptual components and/or robust N400 effects, whereas others showed generally reduced amplitudes. These differences could result from individual variability in the number of trials, the amount of endogenous neural noise, or atypical neural responses in general.

This work is a first step in demonstrating the utility of these implicit measures in assessing vocabulary knowledge in LFAs. Individual variations in which measures best distinguish

'known' and 'unknown' words should be considered in future research using such techniques. One valuable utility of these measures is in by-subject and by-item assessments of which words an individual does or does not know. Our results suggest that before conducting such assessments, pilot testing should be performed with a range of variables to determine which best predict 'known' versus 'unknown' vocabulary. Sets of words that the participant definitely knows and does not know can be used to establish feasibility before extending assessments to target words. Modifications and pilot testing must be performed for each individual.

Although this work focused on implicit measures, participants were allowed to make behavioral responses to maintain attention. Behavioral analyses demonstrated that only WF, SE, and PB understood the visual world task instructions, showing higher accuracy and faster RTs for 'known' than 'unknown' words. These participants also showed some of the largest differences between 'known' and 'unknown' words in the EM measures. Although the behavioral data of DL and HD were less reliable, some EM variables (e.g. *percentage of trials last fixated*) trended in the expected direction for all participants, even if not always statistically significant. In the ERP data, only PB made behavioral responses; yet WF also showed an N400 effect in the absence of a behavioral task. These findings demonstrate the inherent advantage of implicit measures, which need not be restricted to those able to make explicit (or correct) responses.

Difficulties in cognitive testing with LFAs have contributed to some limitations in the current study. Challenges to eye-tracking and EEG data collection, such as movement artifacts, are heightened. This may require modifications to testing protocols to ensure participant comfort and engagement, and/or to data cleaning procedures to ensure maximum data retention. Our EM measures proved particularly sensitive to motion, as reflected in the low data retention rate. For

EEG data, some individuals lost a large percentage of trials despite our extensive cleaning and preprocessing procedure. These challenges may also require multiple testing sessions to collect enough usable data. Four of our five participants performed two or more sessions. This repetition may have influenced the data; for example, the N400 effect is sensitive to repetition (Kutas & Federmeier, 2011). However, no participant saw the same stimulus more than three times, and multiple sessions were performed at least one month apart. Given the importance of collecting enough usable trials, the need for multiple sessions outweighed the potential repetition effects. Nevertheless, this factor should be considered in future studies using similar paradigms.

Despite these challenges, any gains in our currently limited understanding of language comprehension in LFAs far outweigh the difficulties of testing these individuals. Our focus on lower-functioning individuals provides important information about a population that is woefully under-represented in the autism literature. This work also demonstrates the importance of using case-study approaches with low-functioning populations and the utility of single-subject analyses for establishing implicit assessments in individual participants.

These results have important implications for clinicians working with LFAs. Use of these implicit measures for item-by-item identification of 'known' and 'unknown' words could facilitate targeted therapeutic approaches. For example, knowing which words an individual comprehends would allow a language therapist to focus instruction on less understood words, thereby maximizing the use of clinical time and minimizing patient boredom and disengagement. Similarly, the more parents and caregivers know about an individual's comprehension abilities, the more successful their daily interactions and communication will be. Implicit measurement techniques could provide such information and hold far-reaching implications for caring for and treating individuals with autism, especially those without functional speech.

In conclusion, we demonstrate that EMs, PD, and ERPs can provide implicit estimates of receptive vocabulary knowledge in LFAs, although participants differ in individual sensitivity to specific measures. This variability highlights the importance of tailoring these assessments to each individual. Despite the inevitable heterogeneity of our limited number of participants, this work is one of the only studies to use sophisticated neuropsychological methodologies, such as EEG and eye-tracking, to examine language processing in individuals with severe autism, thereby offering a rare insight into this population. The findings have important implications for the development of implicit language assessments in populations unable to provide behavioral responses.

Acknowledgements

The authors would like to thank Ishanti Gangopadhyay for her help with data collection. We are deeply grateful to several anonymous reviewers for their appreciation of the challenges of working with this population and for their thoughtful insights and recommendations. We would especially like to thank all of the participants and their families and/or caregivers and the Linwood Center of Ellicott City, Maryland, an approved IRB research site instrumental in research involving individuals on the autism spectrum.

This research was supported by a grant from the Nancy Lurie Marks Foundation; the Department of Defense Autism Research Program AR093137; The Therapeutic Cognitive Neuroscience Fund; and the Benjamin and Adith Miller Family Endowment on Aging, Alzheimer's, and Autism Research.

References

Anderson, C. J., & Colombo, J. (2009). Larger tonic pupil size in young children with autism spectrum disorder. *Developmental Psychobiology*, *51*(2), 207–211.

Bavin, E. L., Kidd, E., Prendergast, L., Baker, E., Dissanayake, C., & Prior, M. (2014). Severity of Autism is Related to Children's Language Processing. *Autism Research*.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (2nd ed., pp. 142–162). New York, NY: Cambridge University Press.

Brenner, L. A., Turner, K. C., & Müller, R.-A. (2007). Eye movement and visual search: are there elementary abnormalities in autism? *Journal of Autism and Developmental Disorders*, *37*(7), 1289–1309.

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*(3), 896–904.

Byrne, J. M., Dywan, C. A., & Connolly, J. F. (1995). Assessment of children's receptive vocabulary using event-related brain potentials: Development of a clinically valid test. *Child Neuropsychology1*, *1*(3), 221–223.

Center for Disease Control and Prevention, C. (2014). Prevalence of autism spectrum disorder among children aged 8 years - Autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR*, *63*(2), 1–21.

Connolly, J. F., & D'Arcy, R. C. (2000). Innovations in neuropsychological assessment using event-related brain potentials. *International Journal of Psychophysiology*, *37*, 31–47.

Davies, M. (2008). *The Corpus of Contemporary American English: 450 million words, 1990-*

present. Retrieved from http://corpus.byu.edu/coca/

Davis, G., & Plaisted-Grant, K. (2015). Low endogenous neural noise in autism. *Autism*, *19*(3), 351–362.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Delorme, A., Sejnowski, T. J., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*(4), 1443–1449.

Dunn, L., & Dunn, D. (2007). Peabody Picture Vocabulary Tests (4th edition). *American Guidance Service*. Circle Pines, MN: American Guidance Service.

Dunn, M. A., Gaughan Jr, H., Kreuzer, J., & Kurtzberg, D. (1999). Electrophysiologic correlates of semantic classification in autistic and normal children. *Developmental Neuropsychology*, *16*(1), 79–99.

Goldberg, M. C., Lasker, A. G., Zee, D. S., Garth, E., Tien, A., & Landa, R. J. (2002). Deficits in the initiation of eye movements in the absence of a visual target in adolescents with high functioning autism. *Neuropsychologia*, *40*(12), 2039–2049.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461.

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage*, *45*(4), 1199–1211.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related

brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Kaufman, A., & Kaufman, N. (2004). Kaufman Brief Intelligence Test (2nd edition). Circle Pines, MN: American Guidance Service.

Kuipers, J.-R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *5*(61), 1–5.

Kuipers, J.-R., & Thierry, G. (2013). ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex*, *49*(10), 2853–2860.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & Hillyard, S. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., van Petten, C. K., & Kluender, R. (2006). Psycholinguistics Electrified II (1994-2005). In M. Traxler & M. A. Gernsbacher (Eds.), *Handboook of Psycholinguistics2* (2nd editio, pp. 659–724). Academic Press.

Kylliäinen, A., Jones, E. J. H., Gomot, M., Warreyn, P., & Falck-Ytter, T. (2014). Practical Guidelines for Studying Young Children With Autism Spectrum Disorder in Psychophysiological Experiments. *Review Journal of Autism and Developmental Disorders*.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing

the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933.

Ledoux, K., Coderre, E. L., Bosley, L., Buz, E., Gangopadhyay, I., & Gordon, B. (2016). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavior Research Methods*, *48*(1), 285–305.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part 1): Modules 1-4. Torrance, CA: Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685.

Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology (2nd ed)*. London, UK: Chapman & Hall.

Martineau, J., Hernandez, N., Hiebel, L., Roché, L., Metzger, A., & Bonnet-Brilhault, F. (2011). Can pupil size and pupil responses during visual scanning contribute to the diagnosis of autism spectrum disorder in children? *Journal of Psychiatric Research*, *45*(8), 1077–1082.

McCleery, J. P., Ceponiene, R., Burner, K. M., Townsend, J., Kinnear, M., & Schreibman, L. (2010). Neural correlates of verbal and nonverbal semantic integration in children with autism

spectrum disorders. *Journal of Child Psychology and Psychiatry*, *51*(3), 277–286.

Mottron, L., Mineau, S., Martel, G., Bernier, C. S.-C., Berthiaume, C., Dawson, M., … Faubert, J. (2007). Lateral glances toward moving stimuli among young children with autism: Early regulation of locally oriented perception? *Development and Psychopathology*, *19*, 23–36.

Odekar, A., Hallowell, B., Kruse, H., Moates, D., & Lee, C.-Y. (2009). Validity of eye movement methods and indices for capturing semantic (associative) priming effects. *Journal of Speech, Language, and Hearing Research*, *52*, 31–48.

Pérez Velázquez, J. L., & Galán, R. F. (2013). Information gain in the brain's resting state: A new perspective on autism. *Frontiers in Neuroinformatics*, *7*(37), 1–10.

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*(10), 2940–2951.

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2015). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, 1–14. http://doi.org/10.1177/1362361315600146

Schmitt, L. M., Cook, E. H., Sweeney, J. A., & Mosconi, M. W. (2014). Saccadic eye movement abnormalities in autism spectrum disorder indicate dysfunctions in cerebellum and brainstem. *Molecular Autism*, *5*(47), 1–13.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493.

Tanenhaus, M., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations

and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Turner, L. M., Stone, W. L., Pozdol, S. L., & Coonrod, E. E. (2006). Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism*, *10*(3), 243–265.

Footnotes

[1] According to the Institutional Review Board at Johns Hopkins University School of Medicine, a case study constitutes three or fewer participants. As the current study investigates five participants, it is considered "research" and is subject to HIPAA privacy restrictions. As an individualized approach is important in consideration of LFAs, we adhere as closely as possible to a case study-type approach to describe participants and results.

[2] Although we were interested in the implicit measures and did not require behavioral responses, all participants (based on prior experience with computer paradigms) spontaneously sought a task or demonstrated desire to have a task to complete (see Task Procedure for details). We report behavioral data analyses in the Results, though these behavioral responses are not the focus of the current study.

[3] Although PB's scores on the K-BIT and PPVT did not indicate intellectual impairment, he is unable to function in society without assistance due to restricted and repetitive behaviors and deficits in social communication. Therefore, as discussed in the Participants section, he was classified as "low-functioning" for current purposes.

[4] More information is available in Ledoux et al. (2016) about the frequency norming supporting these 'known' and 'unknown' categories in TD adults. Such norming is virtually impossible with LFAs for the very reason that we sought to use implicit measures of assessment: their verbal and other behavioral responses are extremely variable and often unreliable.

[5] Some variables have different degrees of freedom because different numbers of trials went into the analyses. On some trials the participant did not look at the target picture at all. In such cases, the *mean fixation duration* would have a value of 0 and would be included in the analysis, but *first* fixation duration would be coded as "not applicable" (NA) and would not be included.

Tables

Table 1: Participant demographics, including autism diagnostic test results (ADOS, ADI-R), intelligence scores (K-BIT), and vocabulary scores (PPVT).

Participant (not their real initials)	Age	Gender	ADOS				ADI-R				K-BIT		PPVT
			Version	Module	Social + communication total	Classification	Social interaction	Communication	Behaviors	Development	Verbal	Non-verbal	
DL	18	M	1	1 (adapted)	20	autism	22	14	2	5	Not available		Not available
HD	15	M	Not available				22	20	6	5	Not available		20
WF	39	M	2	4 (adapted)	22	autism	Not available				45	79	58
SE	40	M	2	2	20	autism	Not available				40	60	43
PB	48	M	2	2	19	autism	Not available				93	131	94

Table 2: Total number of collected and usable trials for each participant. Note that for the EM and PD data, a full session was 160 trials; for the EEG data, a full session was 320 trials.

Participant	**Eye Movement (EM)**		**Pupillometry (PD)**		**Event-Related Potentials (ERP)**	
	Total number of trials recorded	**Total number of usable trials**	**Total number of trials recorded**	**Total number of usable trials**	**Total number of trials recorded**	**Total number of usable trials**
DL	120	62	120	56	780	362
HD	160	67	160	107	320	203
WF	160	134	160	155	640	462
SE	320	93	320	174	640	289
PB	160	114	160	122	640	375

Table 3: Individual behavioral data for the visual world and picture-word congruity tasks. For reaction times, standard error of the mean (SE) is shown in parentheses.

Participant	Visual world task				Picture-word congruity task			
	Accuracy (%)		Reaction time (ms)		Accuracy (%)		Reaction time (ms)	
	'known'	*'unknown'*	*'known'*	*'unknown'*	*'known'*	*'unknown'*	*'known'*	*'unknown'*
DL	38	28	1529 (121)	1638 (138)	Did not provide behavioral responses			
HD	59	27	2618 (166)	3003 (151)	Did not provide behavioral responses			
WF	100	49	2583 (93)	3831 (101)	Did not provide behavioral responses			
SE	98	28	1432 (53)	1710 (99)	Did not provide behavioral responses			
PB	100	64	1343 (56)	3151 (134)	Not enough reliable data available for analysis			

Table 4: Summary table of results for each participant.

Measures	Participant				
	DL	**HD**	**WF**	**SE**	**PB**
Behavioral: visual world task	No differences between 'known' and 'unknown' trials in accuracy or RT	No differences between 'known' and 'unknown' trials in accuracy or RT	Higher accuracy and shorter RTs for 'known' word trials	Higher accuracy and shorter RTs for 'known' word trials	Higher accuracy and shorter RTs for 'known' word trials
Behavioral: picture-word congruity task	Could not understand task, did not provide behavioral responses	Could not understand task, did not provide behavioral responses	Could not understand task, did not provide behavioral responses	Could not understand task, did not provide behavioral responses	Available behavioral data not reliable, not analyzed
EM	No statistically significant effects	No statistically significant effects	Significant effects for *mean fixation duration, first fixation duration, first dwell, percent fixation duration, percent last fixated, and number of fixations*	Significant effect for *percent last fixated*	Significant effects for *percent fixation duration on stimulus, percent dwell, percent last fixated, number of fixations*
PD	No statistically significant effects	No statistically significant effects	No statistically significant effects	No statistically significant effects	Significant effects for *peak dilation, max percent dilation*
ERP	Difference between conditions for 'unknown' words only, likely due to noise	No statistically significant N400 effects	Significant N400 effect for 'known' words at Pz cluster from 200-400 ms	No statistically significant N400 effects	Significant N400 effect for 'known' words at C3 cluster from 400-550 ms

Figure Captions

Figure 1: Examples of ‘known’ and ‘unknown’ stimuli.

Figure 2: Illustration of the nine electrode clusters used for EEG analysis.

Figure 3: Results for DL. a) Bar graphs comparing ‘known’ and ‘unknown’ word trials for each of the EM variables. b) Comparisons of ‘known’ and ‘unknown’ word trials for each of the pupillometry variables. c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up. The grey bar beneath the waveforms indicates significant differences between congruent and incongruent conditions for ‘unknown’ words, as determined by permutation tests with a cluster-based FWE correction at $p < .05$.

Figure 4: Results for HD. a) Bar graphs comparing ‘known’ and ‘unknown’ word trials for each of the EM variables. b) Comparisons of ‘known’ and ‘unknown’ word trials for each of the pupillometry variables. c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up.

Figure 5: Results for WF. a) Bar graphs comparing ‘known’ and ‘unknown’ word trials for each of the EM variables. b) Comparisons of ‘known’ and ‘unknown’ word trials for each of the pupillometry variables. Significant differences between ‘known’ and ‘unknown’ words, based on permutation tests with Bonferroni corrections, are indicated by asterisks (** = $p < .01$; * = $p < .05$). c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up. The orange bar beneath the waveforms indicates significant differences between congruent and incongruent conditions for ‘known’ words, as determined by permutation tests with a cluster-based FWE correction at $p < .05$.

Figure 6: Results for SE. a) Bar graphs comparing ‘known’ and ‘unknown’ word trials for each of the EM variables. b) Comparisons of ‘known’ and ‘unknown’ word trials for each of the pupillometry variables. Significant differences between ‘known’ and ‘unknown’ words, based on permutation tests with Bonferroni corrections, are indicated by asterisks * = $p < .05$). c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up.

Figure 7: Results for PB. a) Bar graphs comparing ‘known’ and ‘unknown’ trials for each of the EM variables. b) Comparisons of ‘known’ and ‘unknown’ trials for each of the pupillometry variables. Significant differences or trends toward significance between ‘known’ and ‘unknown’ words, based on permutation tests with Bonferroni corrections, are indicated by asterisks (** = $p < .01$; * = $p < .05$; ‡ = $p < .10$). c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up. The orange bar beneath the waveforms indicates significant differences between congruent and incongruent conditions for ‘known’ words, as determined by permutation tests with a cluster-based FWE correction at $p < .05$.

Figure 8: Summary descriptions of individual subject data. a) ‘unknown’-’known’ difference scores (scaled *z*-scores) for each EM and PD variable. Variables on the left were predicted to be larger for ‘unknown’ than ‘known’ word trials, so the ‘unknown’-’known’ difference score should be negative. Variables on the right were predicted to be larger for ‘known’ than ‘unknown’ word trials, so ‘unknown’-’known’ differences should be positive. b) Topographic plots of the ERP incongruent-congruent difference for ‘known’ and ‘unknown’ words in 50 ms windows from 200 to 800 ms after sound presentation.

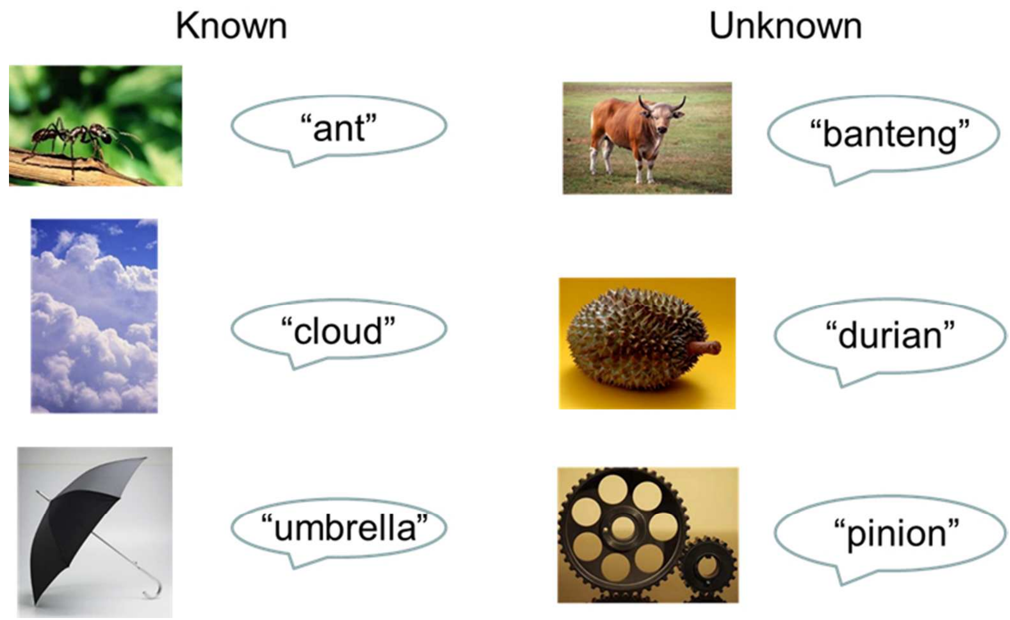


Figure 1: Examples of ‘known’ and ‘unknown’ stimuli.

290x182mm (72 x 72 DPI)

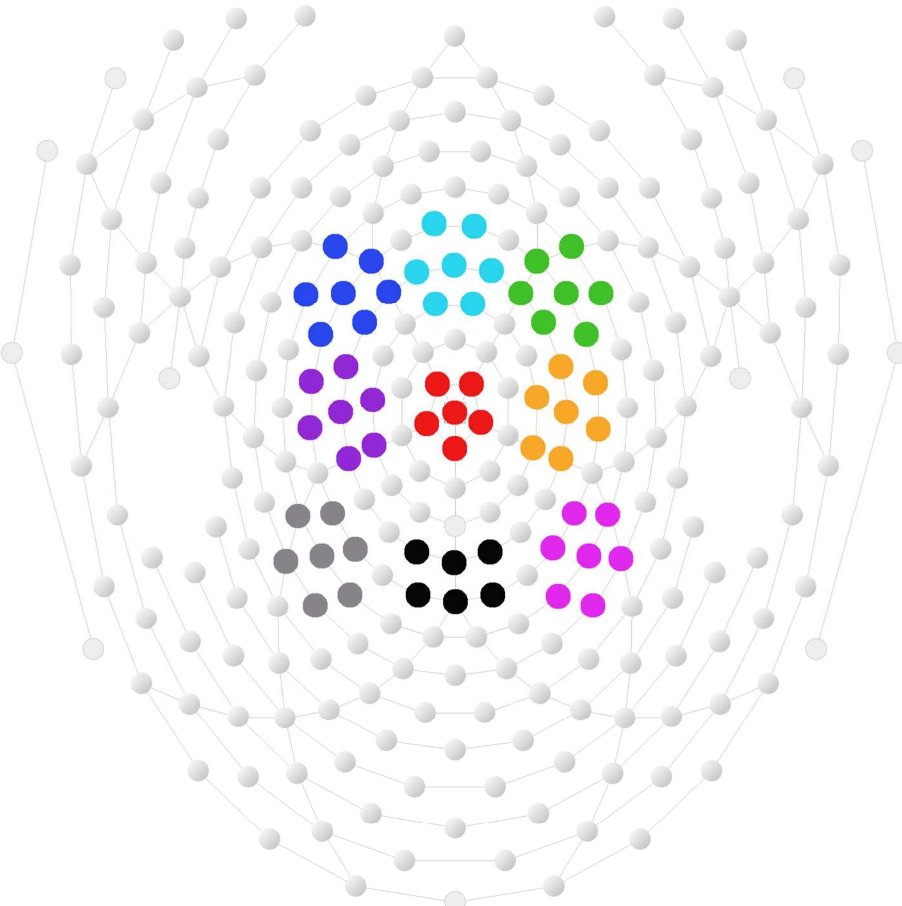


Figure 2: Illustration of the nine electrode clusters used for EEG analysis.

529x493mm (72 x 72 DPI)

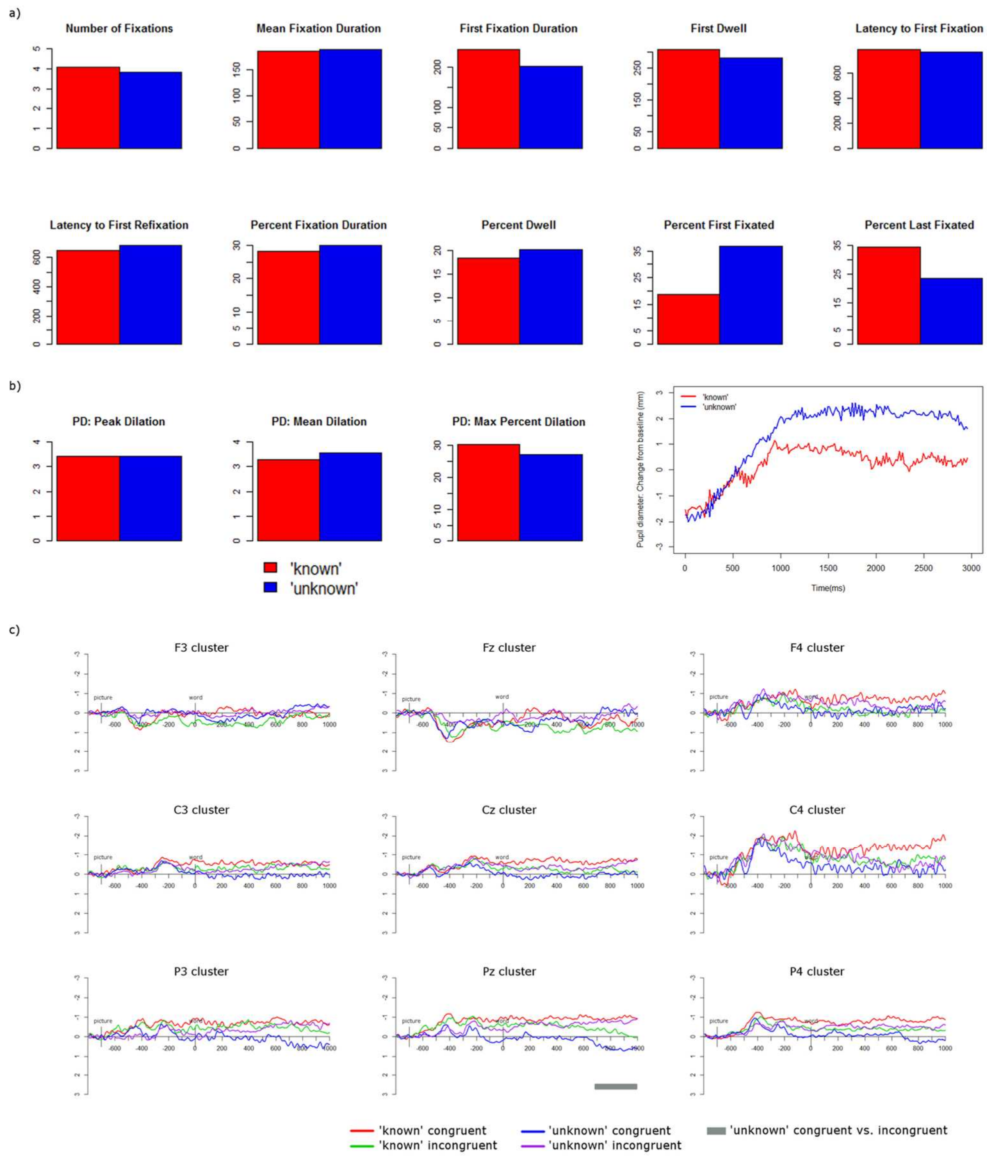


Figure 3: Results for DL. a) Bar graphs comparing 'known' and 'unknown' word trials for each of the EM variables. b) Comparisons of 'known' and 'unknown' word trials for each of the pupillometry variables. c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up. The grey bar beneath the waveforms indicates significant differences between congruent and incongruent conditions for 'unknown' words, as determined by permutation tests with a cluster-based FWE correction at $p < .05$.

127x146mm (300 x 300 DPI)

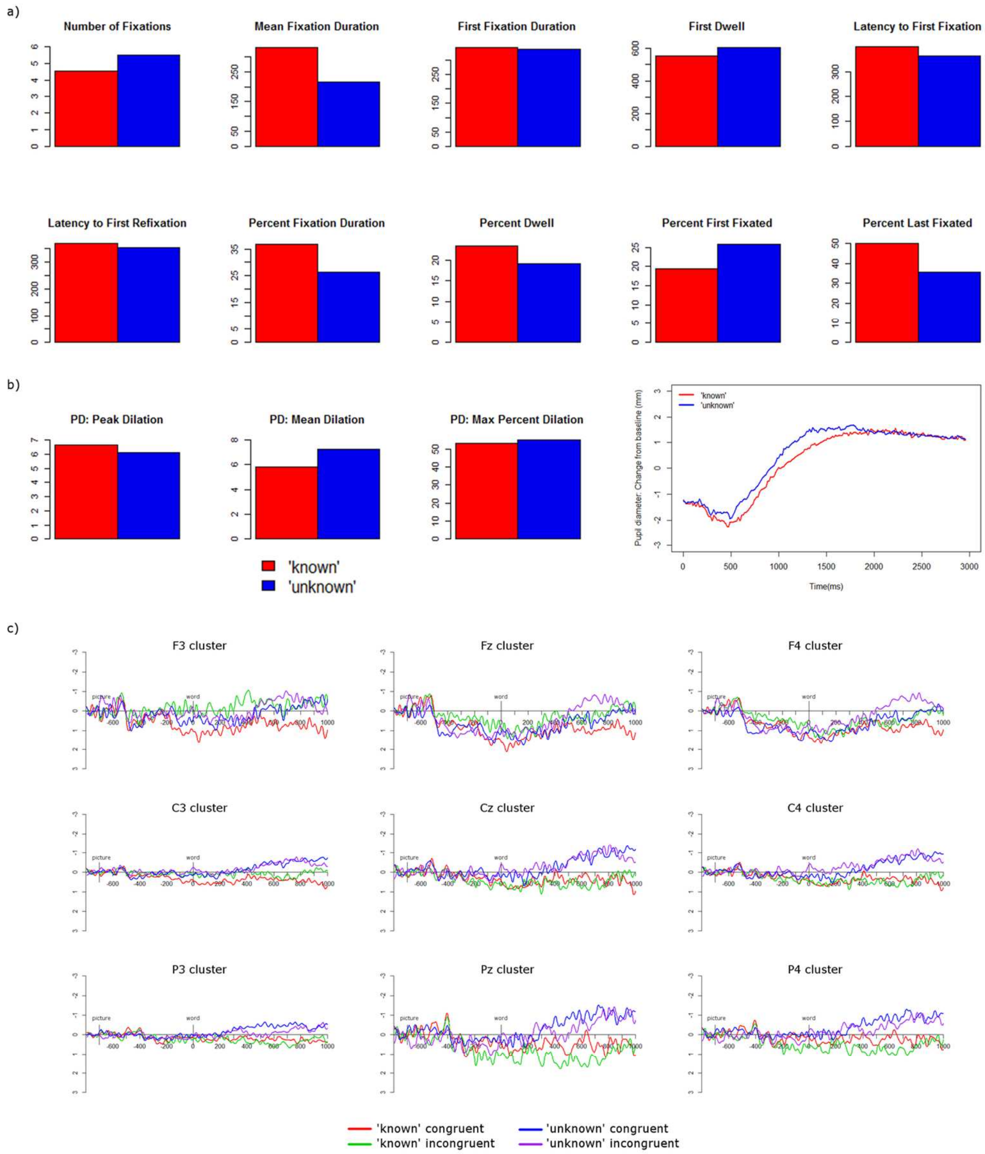


Figure 4: Results for HD. a) Bar graphs comparing 'known' and 'unknown' word trials for each of the EM variables. b) Comparisons of 'known' and 'unknown' word trials for each of the pupillometry variables. c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up.

127x146mm (300 x 300 DPI)

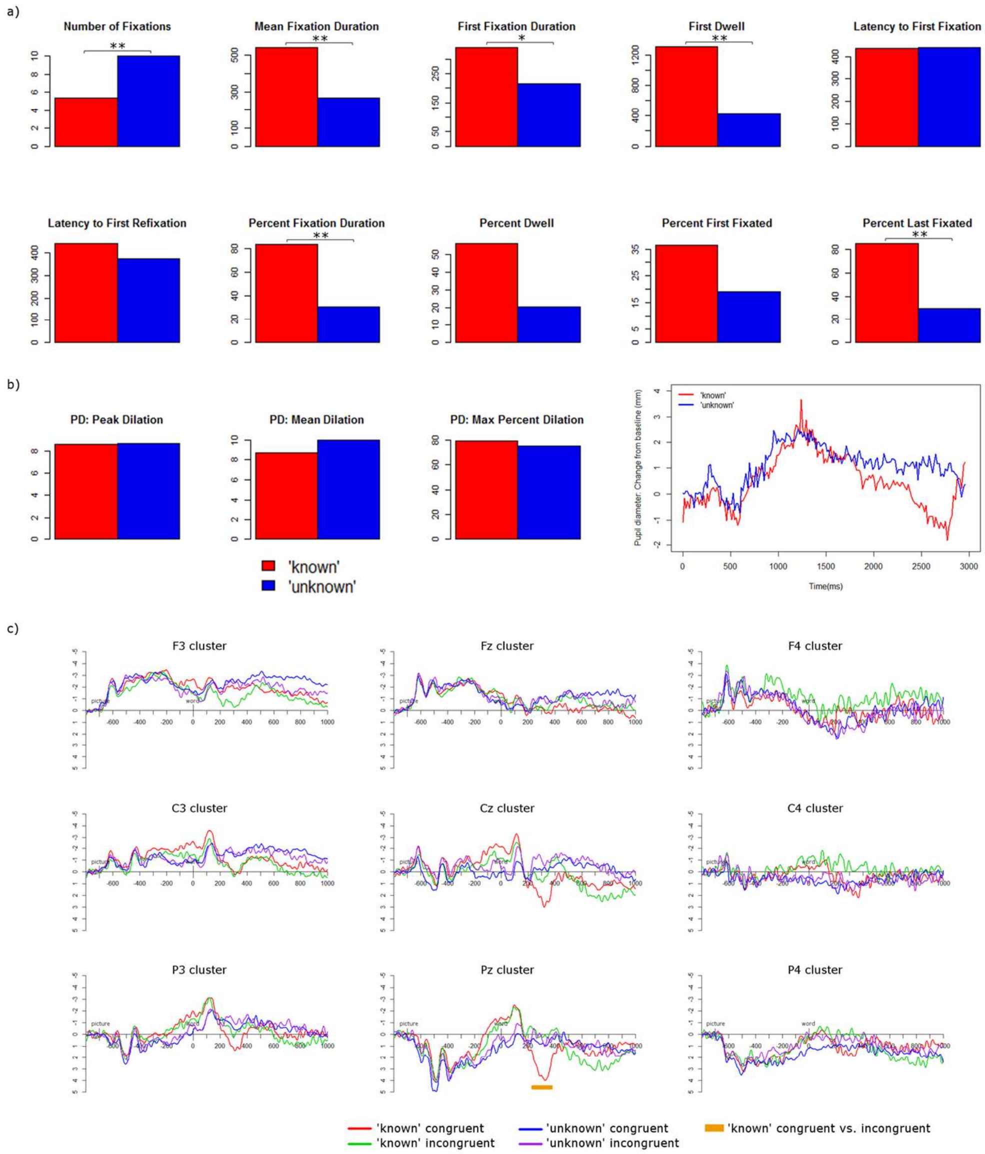


Figure 5: Results for WF. a) Bar graphs comparing 'known' and 'unknown' word trials for each of the EM variables. b) Comparisons of 'known' and 'unknown' word trials for each of the pupillometry variables. Significant differences between 'known' and 'unknown' words, based on permutation tests with Bonferroni corrections, are indicated by asterisks (** = p < .01; * = p < .05). c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up. The orange bar beneath the waveforms indicates significant differences between congruent and incongruent conditions for 'known' words, as determined by permutation tests with a cluster-based FWE correction at p < .05.

127x146mm (300 x 300 DPI)

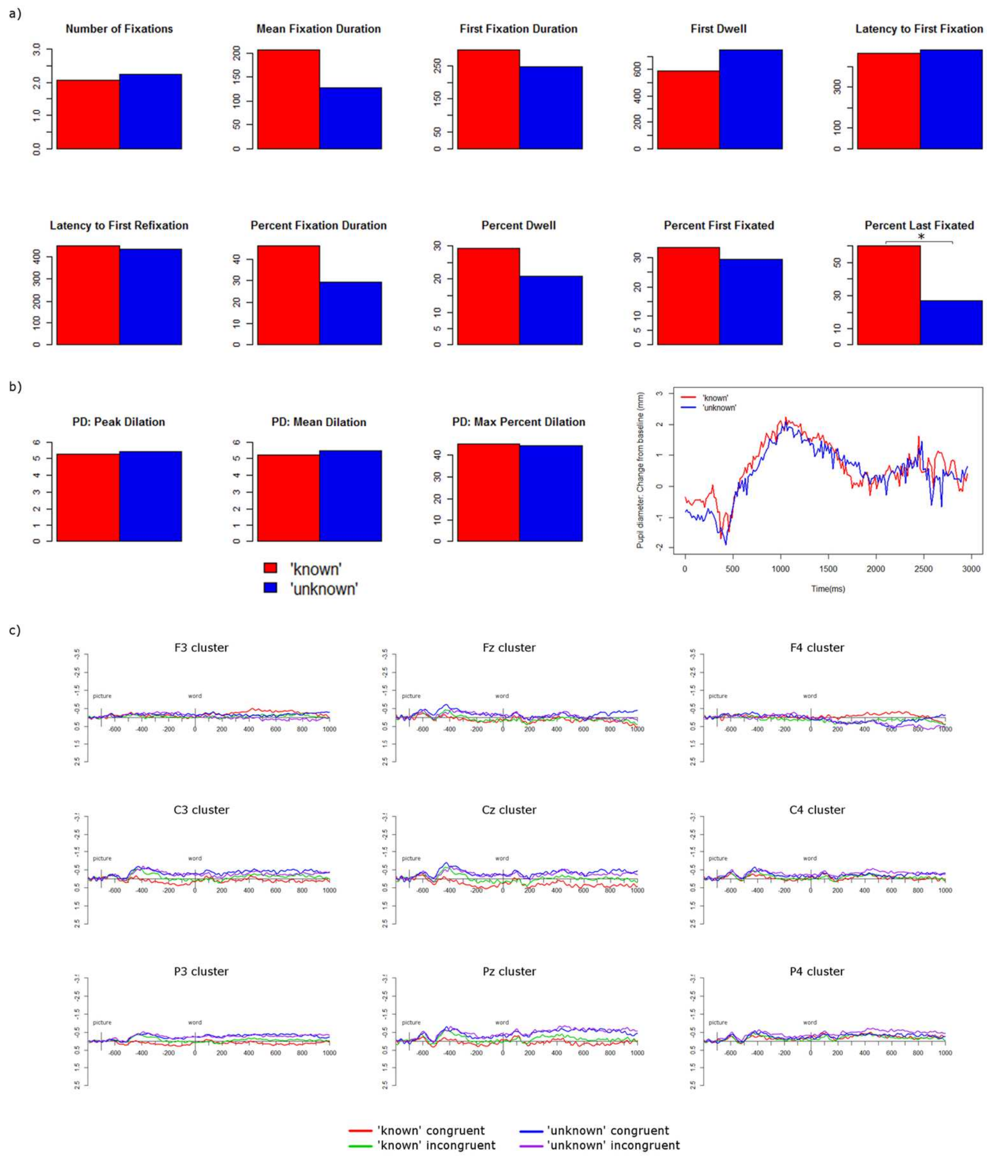


Figure 6: Results for SE. a) Bar graphs comparing 'known' and 'unknown' word trials for each of the EM variables. b) Comparisons of 'known' and 'unknown' word trials for each of the pupillometry variables. Significant differences between 'known' and 'unknown' words, based on permutation tests with Bonferroni corrections, are indicated by asterisks * = p < .05). c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up.

127x146mm (300 x 300 DPI)

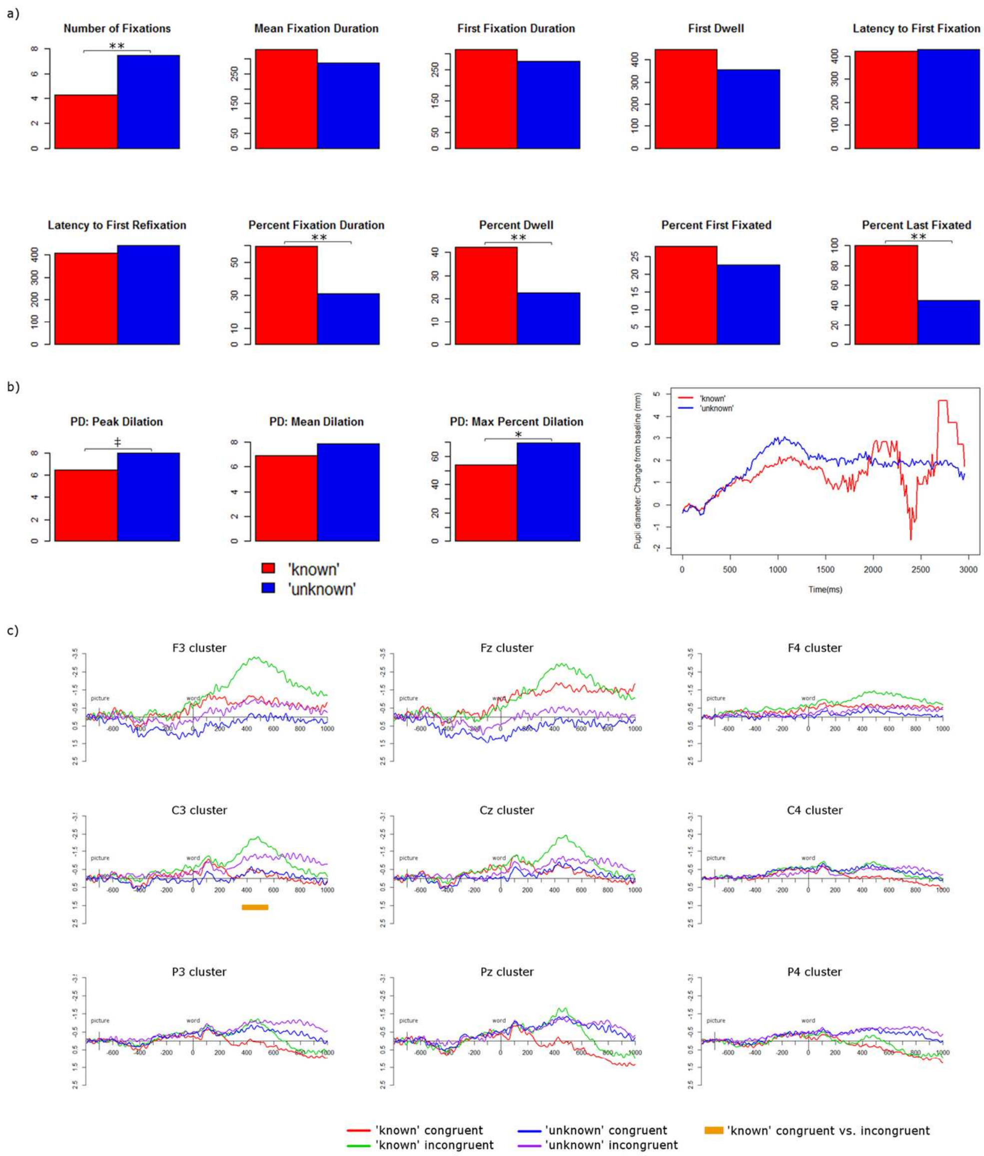


Figure 7: Results for PB. a) Bar graphs comparing 'known' and 'unknown' trials for each of the EM variables. b) Comparisons of 'known' and 'unknown' trials for each of the pupillometry variables. Significant differences or trends toward significance between 'known' and 'unknown' words, based on permutation tests with Bonferroni corrections, are indicated by asterisks (** = p < .01; * = p < .05; ‡ = p < .10). c) ERP data for all conditions at the 9 electrode cluster sites. Negative is plotted up. The orange bar beneath the waveforms indicates significant differences between congruent and incongruent conditions for 'known' words, as determined by permutation tests with a cluster-based FWE correction at p < .05.

127x146mm (300 x 300 DPI)

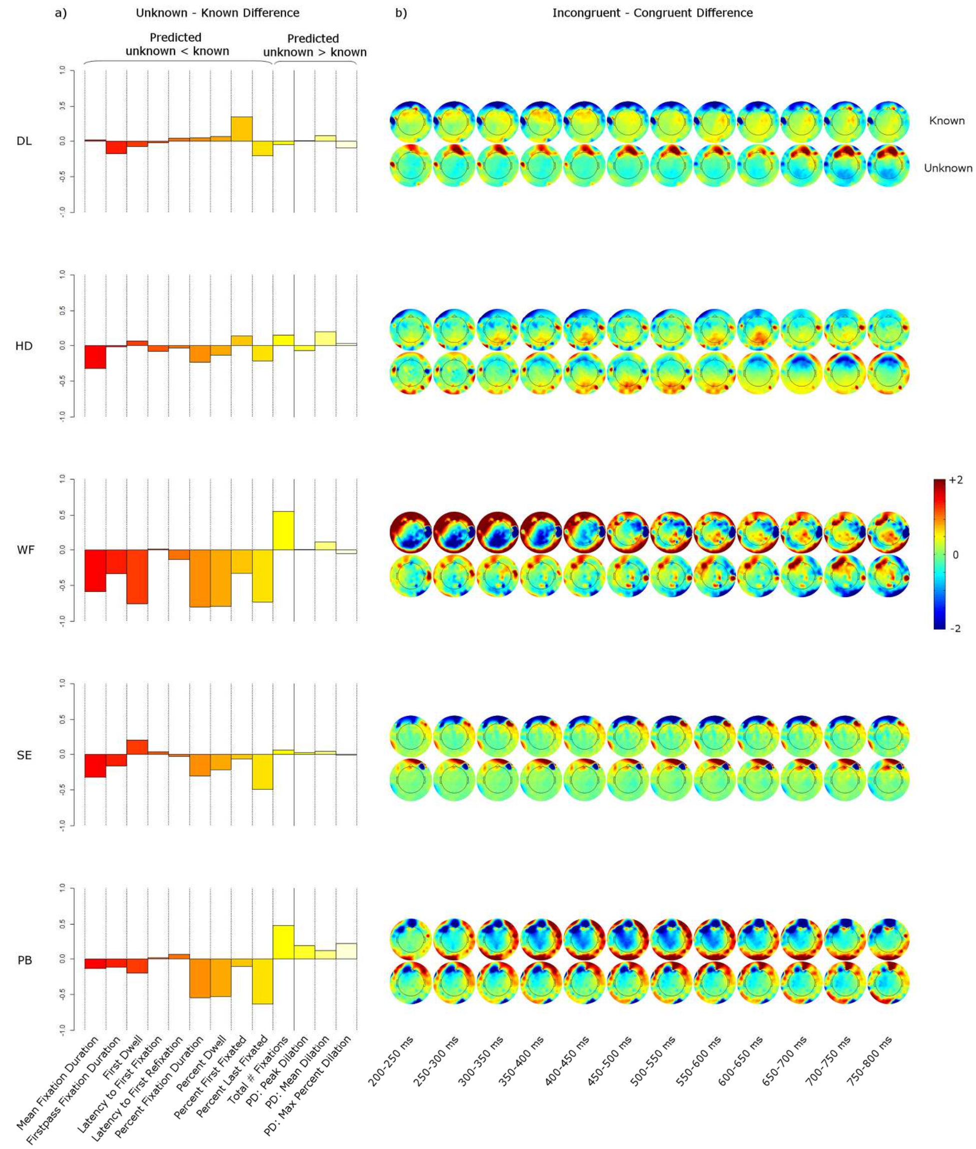


Figure 8: Summary descriptions of individual subject data. a) 'unknown'-'known' difference scores (scaled z-scores) for each EM and PD variable. Variables on the left were predicted to be larger for 'unknown' than 'known' word trials, so the 'unknown'-'known' difference score should be negative. Variables on the right were predicted to be larger for 'known' than 'unknown' word trials, so 'unknown'-'known' differences should be positive. b) Topographic plots of the ERP incongruent-congruent difference for 'known' and 'unknown' words in 50 ms windows from 200 to 800 ms after sound presentation.

211x241mm (300 x 300 DPI)

Supplementary Material

Supplementary Material 1: Picture stimuli for each of the ‘known’ and ‘unknown’ words used in the experiments

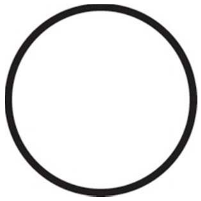
‘Known’		**‘Unknown’**	
Word	**Picture**	**Word**	**Picture**
airplane		ablution	
ant		acerola	
apple		ackee	
baby		addax	
ball		agouti	

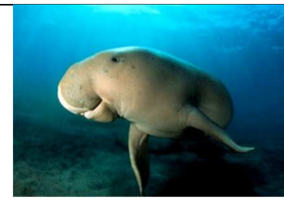
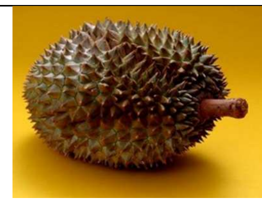
balloon		anemometer	
banana		angklung	
bathtub		anole	
bed		argali	
bicycle		avocet	
book		babirusa	

boots		balalaika	
bottle		banteng	
bowl		barasingha	
box		bilby	
boy		binturong	

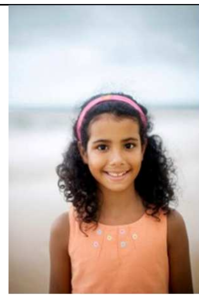
bread		bolster	
brush		caiman	
bus		cainito	
butterfly		capybara	
cake		caracal	
camera		carambola	

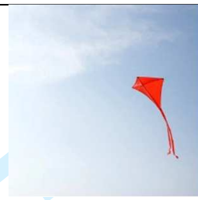
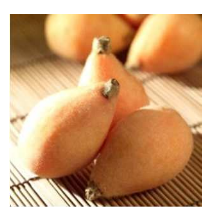
candy		carboy	
car		celeriac	
cat		chayote	
chair		cherimoya	
cheese		civet	

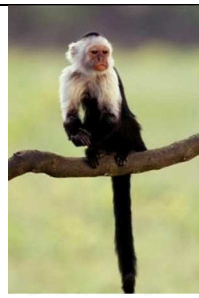
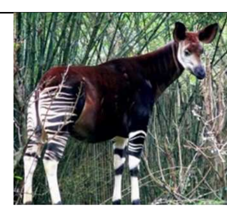
circle		colugo	
clock		conflagration	
cloud		confluence	
coat		cudgel	
cookie		douc	

cow		drachma	
crayons		dugong	
cup		durian	
dinosaur		echidna	
dog		effigy	
door		epee	

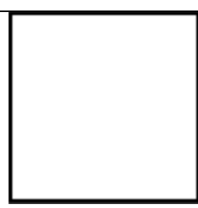
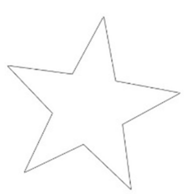
drum		feijoa	
elephant		floe	
flower		fossa	
fork		frieze	
frog		gelada	

girl		gerenuk	
grapes		greengage	
hammer		harrow	
horse		homogenizer	
house		jerboa	

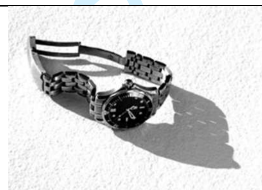
key		jicama	
kite		jujube	
knife		kinkajou	
ladder		kohlrabi	
leaf		kumquat	
lion		loquat	

monkey		mead	
mouse		medlar	
orange		melee	
pencil		mendicant	
pig		millet	
pot		okapi	

pretzel		pangolin	
rabbit		panoply	
scissors		peccary	
shoes		persimmon	
slide		pillory	
snake		pinion	

snowman		quince	
sock		raiment	
spider		ramekin	
spoon		repast	
square		rowan	
star		saguaro	

swing		specie	
table		sylph	
telephone		talisman	
tiger		tamarillo	
train		tamarind	

tree		tarsier	
umbrella		visage	
watch		yangmei	

Appendix 3

Coderre, E., Gordon, B., & Ledoux, K. (Under revision.) The use of mixed-effects models to predict receptive vocabulary knowledge from implicit measures of language comprehension.

THE USE OF MIXED-EFFECTS MODELS TO PREDICT RECEPTIVE VOCABULARY KNOWLEDGE FROM IMPLICIT MEASURES OF LANGUAGE COMPREHENSION

Emily L. Coderre[1*], Barry Gordon[1,2], & Kerry Ledoux[1]

[1]Cogntivity Neurology/Neuropsychology Division, Department of Neurology, The Johns Hopkins University School of Medicine; Baltimore, Maryland, USA

[2]Department of Cognitive Science
The Johns Hopkins University; Baltimore, Maryland, USA

*Address for Correspondence (EC):
Emily Coderre
Johns Hopkins University School of Medicine
Cognitive Neurology/Neuropsychology-Department of Neurology
1629 Thames Street, Suite 350
Baltimore, MD 21231
Phone: 001-443-287-1716
Fax: 001-410-955-0188
Email: ecoderr1@jhmi.edu, emily.coderre@gmail.com

ABSTRACT

The cognitive operations underlying language are frequently assessed using overt behaviors like reaction time or verbal report. However, such measures assume an understanding of task goals and an ability to execute the required response. In certain populations, such as low-functioning non-verbal individuals with autism, these measures might be difficult or impossible to obtain, making implicit measures of cognition essential. In recent work, we have shown that eye movements (EMs), pupillary dilation (PD), and event-related potentials (ERPs) can be used as implicit measures of vocabulary knowledge both in normal adults (Ledoux et al., 2015) and in low-functioning individuals with autism (Coderre et al., submitted). During a forced-choice recognition task (EM and PD) and a picture-word congruity task (ERP), objectively-classified “known” (high-frequency) and “unknown” (low-frequency) words showed consistent differences across all three implicit measures. The utility of these implicit measures in distinguishing between known and unknown vocabulary holds the potential to be able to predict, on an item-level basis, whether an individual knew a particular word or not based on their patterns of EM, PD, and ERPs in response to that word. This predictive ability would be extremely valuable for individuals who are unable to give an overt behavioral report of their knowledge ratings, such as some low-functioning individuals with autism. The aim of the current work is to demonstrate how regression modeling can be used to estimate the latent variable of receptive vocabulary knowledge from implicit measures (eye movements, pupillary dilation, and event-related potentials) to allow for the measurement of knowledge even in the absence of an overt behavioral response. A linear mixed effects model was trained on data from normal adults. Subjective knowledge ratings of each word, provided after experiment completion, served as the dependent variable while 13 measures taken from the EM, PD, and ERP data served as independent variables. Cross-validation demonstrated that the model was able to predict subjective knowledge ratings from implicit measures with a high accuracy rate. Implicit EM, PD, and ERP measures from five low-functioning individuals with autism were then entered into the previously-built model to predict subjective knowledge for each word. Overall, this work suggests that regression modeling can predict receptive vocabulary knowledge in the absence of behavioral responses. Such a technique holds important implications for assessment of language comprehension in populations for whom explicit behavioral responses might be difficult or impossible to obtain and offers the potential for extension to other aspects of cognition such as memory, consciousness, or reasoning.

Keywords: regression modeling, implicit measures, vocabulary knowledge, event-related potentials, eye movements, pupil dilation

THE USE OF MIXED-EFFECTS MODELING TO PREDICT RECEPTIVE VOCABULARY KNOWLEDGE FROM IMPLICIT MEASURES OF LANGUAGE COMPREHENSION

1. Introduction

One challenge in the study of cognition is to gain access to the mental representations and processes that are at the heart of language, memory, and thought. Because direct access to or observation of latent constructs of interest is impossible, researchers have long inferred the operation and quantification of such latent variables through more overt, observable behaviors, such as the time taken to respond to a stimulus or a participant's verbal report of their mental experience. Such overt or explicit measures, while extremely valuable, are subject to multiple influences, such as attention and motivation. Additionally, such explicit measures can only be obtained from individuals who are able to execute the requisite response, and may be difficult or impossible to obtain with certain populations who cannot speak or reliably execute complex behaviors (such as infants, nonverbal individuals with autism, or patients in coma).

The development and use of more implicit measures of cognition that do not rely on overt verbal or behavioral responses may allow for an alternative assessment of latent cognitive variables across a wider range of participants. Implicit measurement techniques such as eye movement monitoring, functional neuroimaging, or electroencephalography have enjoyed a popularity of use in recent years for the precise reason that they do not rely on overt behavioral responses, and have led to insights about cognitive processing across a wide range of participant populations. One particular use of these techniques that has been relatively less explored (although certainly not ignored) is the extrapolation from existing implicit measurement data to make predictions about new implicit responses, given the state of the underlying latent construct. Regression modeling is the standard tool for such extrapolations. The aim of the current paper is to use regression modeling to estimate the latent variable of receptive vocabulary knowledge from implicit measures (specifically eye movements, changes in pupillary dilation, and event-related potentials) to allow for the measurement of knowledge even in the absence of an overt behavioral response. After demonstrating the utility of the modeling procedure for predicting vocabulary knowledge from implicit measures, we demonstrate how this methodology can be extended to estimate receptive vocabulary knowledge in a group of low-functioning individuals with autism, who cannot make reliable overt responses but for whom an estimate of vocabulary knowledge would be especially useful from a clinical or educational standpoint.

1.1. *Implicit measures of language processing*

Eye movement monitoring, measures of pupillary dilation, and event-related potentials have all proven useful as implicit measures of language processing. Eye movement (EM) paradigms have long proven useful in the study of reading behavior (Rayner, 1998), and the development of the visual world paradigm has extended the use of this technique to the study of other aspects of language comprehension and production (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Tanenhaus, Magnuson, Dahan, & Chambers, 2000; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In the visual world paradigm, participants typically see a visual

display of pictures followed by a spoken word or phrase; participant's eyes generally move quickly and reliably to a named picture as soon as that named entity can be disambiguated from other pictures in the display. Using a similar paradigm with visual displays of pictures following presentation of a written word, Odekar et al. (2009) observed that eye movements were faster and fixations were longer when a picture was semantically related to the prime word, compared to when the pictures were unrelated to the prime word. Thus eye movement patterns can reflect semantic priming during language comprehension, without relying on an overt behavioral response.

Pupillary dilation (PD) can be measured by time-locking changes in pupil diameter to the onset of an external stimulus. PD increases with task difficulty or processing load, therefore changes in PD have been interpreted as a measure of resource recruitment (Beatty & Lucero-Wagoner, 2000; Granholm, Asarnow, Sarkin, & Dykes, 1996). Within the domain of language, PD has been used as a measure of processing demands in studies of visual letter perception, semantic and syntactic processing, and even simultaneous interpretation (e.g. Hyönä, Tommola, & Alaja, 1995; Schluroff, 1982; see Beatty & Lucero-Wagoner, 2000 for a review). For instance, in semantic priming experiments, PD increases to a greater extent in response to unrelated pairs of pictures and spoken words compared to related pairs, indicating increased cognitive load and greater resource recruitment in the unrelated condition (Kuipers & Thierry, 2011, 2013).

Event-related potentials (ERPs) are time-locked changes in the electroencephalographam (EEG) elicited by a stimulus. Various individual ERP components have been reliably associated with different aspects of language processing (Rugg & Coles, 1995; Sereno & Rayner, 2003). Most important to the current purposes, the N400 ERP component has been associated with semantic processing and integration (Kutas & Hillyard, 1980; Lau, Phillips, & Poeppel, 2008). In particular, reductions in the amplitude of the N400 are observed when a word is more readily integrated with its context (for example, when the word is congruent with the context, or has a higher cloze probability), relative to when semantic integration is more difficult (when a word is incongruent with its context or has a lower cloze probability). The magnitude of the reduction in N400 amplitude is referred to as the "N400 effect". The N400 effect is taken as a measure of semantic integration, with greater resource recruitment required for incongruent conditions. Previous work has shown that an N400 effect is elicited in response to mismatching pairs of pictures and words, even in young children (Friedrich & Friederici, 2004), but only when the word is within an individual's vocabulary range (Connolly & D'Arcy, 2000).

Our group has recently built upon previous work (Connolly & D'Arcy, 2000; Friedrich & Friederici, 2004; Kuipers & Thierry, 2011, 2013; Odekar et al., 2009) by using these three measures concurrently to assess receptive vocabulary knowledge in a group of normal adult participants. Ledoux et al. (2015) presented participants with very high-frequency words such as *bus* and *horse* that were expected to be known to the majority of participants (hereafter termed "known" words) and very low-frequency words such as *ackee* and *cherimoya* that were expected to be unknown to the majority of participants (hereafter termed "unknown" words) in two experimental paradigms. In a visual world paradigm, participants were presented with four pictures on the screen followed by a spoken word that matched one of the pictures; EM and PD data were collected during this task to evaluate eye-gaze patterns as participants searched for the picture that matched the spoken word, and changes in PD following the presentation of a known

or unknown word. In a picture-word congruency paradigm, participants were presented with a picture followed by a spoken word that either matched did not match the picture. ERP data were collected during this task to evaluate the electrophysiological response to stimulus congruity.

In this population of normal adults, Ledoux et al. (2015) observed that all three measures showed reliable differences in the processing of known and unknown words. Specifically, EMs were faster to known than to unknown words; fixations were longer to known than unknown words; and end-of-trial fixations were more often on the correct named picture for known than unknown words. Results from the PD data showed that changes in PD from baseline were larger, reflecting greater resource recruitment, for unknown than for known words. Finally, the ERP data showed an N400 effect for known words, such that the amplitude of the N400 was reduced for congruent word-picture pairings relative to incongruent pairings, but this effect was absent for unknown words, for which participants could not use prior knowledge to ease integration between the two stimuli. Therefore this prior work demonstrated that EM, PD, and ERP measures can be used together to distinguish between known and unknown words in a population of normal adults.

1.2. *Implicit measures of vocabulary in low-functioning autism*

The use of implicit measures of cognition is especially useful for clinical populations who are less able to provide reliable behavioral reports of their knowledge or abilities. One such population is low-functioning individuals with autism. Autism is a pervasive developmental disorder that is characterized by language delay and impairments that are often severe. Approximately 25% of individuals with autism spectrum disorder (ASD) have little to no functional speech and are characterized as “non-verbal” (Turner, Stone, Pozdol, & Coonrod, 2006). Overt verbal or behavioral reports of vocabulary knowledge are difficult to obtain in many of these individuals due to a lack of functional speech and/or difficulties with understanding or following task instructions. However, this does not preclude functional language comprehension. Implicit measures of receptive language abilities may thus hold enormous potential for this population by offering an alternative assessment of vocabulary knowledge and guidance for future training.

In previous work (Coderre et al., submitted) using the same two tasks described above, we have demonstrated that EM, PD, and ERPs can also distinguish between objectively-rated known and unknown words in a population of low-functioning individuals with autism (LFAs). Although there was significant heterogeneity among participants, measures from all three methodologies showed differences between known and unknown words. More specifically, in the EM data the number of total fixations was smaller for known than unknown words; fixations were longer to known than unknown words; and end-of-trial fixations were more often on the correct named picture for known than unknown words. In the PD data, changes in pupillary dilation from baseline were larger for unknown than for known words. In the EEG data, there was a trend of an N400 effect for known words but not for unknown words. Thus despite heterogeneity between LFA participants, overall these implicit measures were able to distinguish between objectively-rated known and unknown vocabulary for LFAs in patterns that were similar to those reported in normal adults by Ledoux et al. (2015). This suggests that the implicit measures of EM, PD, and ERPs can serve as implicit measures of vocabulary knowledge, without reliance on behavioral measures, for both typical and clinical populations.

1.3. *Using regression modeling to predict vocabulary knowledge from implicit measures of cognition*

One limitation in this previous work is that words were objectively classified into known and unknown categories based on their frequency in the English language: high-frequency words were expected to be known by the majority of participants and were therefore deemed "known", whereas low-frequency words were expected to be unfamiliar by the majority of participants and were therefore deemed "unknown". Individual variations in lexical knowledge – i.e., the fact that some participants may have been familiar with some of the "unknown" words or unfamiliar with some of the "known" words – were not taken into account.

However, implicit measures of knowledge may be able to provide finer-grained variations such that one would be able to assess, on an item-level basis, whether an individual knew a particular word or not based on their patterns of EM, PD, and ERPs in response to that word. This predictive ability would be extremely valuable for individuals who are unable to give an overt behavioral report of their knowledge ratings, such as some low-functioning individuals with autism. The aim of the current study is to demonstrate that implicit measures of EM, PD and ERPs can be used to estimate latent vocabulary knowledge through the use of regression modeling.

Regression procedures estimate the parameters that best describe the relationship between an observed dependent variable and one or more observed independent variables. These estimated parameters can then be used with a different set of observed independent variables to predict the unobserved dependent variable for a new sample. In this way, a regression model can be trained on an initial dataset and then used to predict outcomes for a new population.

To estimate the initial model parameters, model training was performed using the implicit data from the normal adults tested in Ledoux et al. (2015). In addition to providing implicit data, these participants also provided subjective knowledge ratings for each word by rating their knowledge of each word they had encountered on scale from "completely unknown" to "completely known". During model training, we fit a regression model that describes the relationship between the observed dependent variable (the subjective knowledge ratings) and the observed independent variables (the EM, PD, and ERP measures).

The ultimate aim of this work is to develop a means of estimating latent vocabulary knowledge that relies only on the implicit data itself, for use even with participants who cannot provide reliable indicators of their own knowledge through behavioral responses. To do so, we used the regression model trained on the normal adult data to predict knowledge ratings for a group of LFA participants (previously tested in Coderre et al., submitted) using only their implicit measures. Given the difficulties with testing low-functioning individuals and the often unreliable nature of their behavioral responses, the ability to predict vocabulary knowledge from implicit measures of language abilities holds enormous potential for the assessment of cognitive abilities in these populations.

2. Methods

Details of the methodology can also be found in Ledoux et al. (2015) and Coderre et al., submitted.

2.1. *Stimuli*

Stimuli were 160 auditory words with matching pictures. Half of the stimuli were very high-frequency words (average SubtlexUS (Brysbaert & New, 2009) log10 frequency rating = 3.14, *SD* = 0.6), such as *telephone*, *door*, and *circle*. Because of their very high frequency, these words were expected to be known by the majority of participants, and were thus objectively classified as "known." The remaining 80 stimuli were very low-frequency words (average SubtlexUS log10 frequency rating = 0.85, *SD* = 0.5), such as *douc*, *melee*, and *conflagration*. These words were objectively classified as "unknown." All words were highly imageable. Although an effort was made to include a range of word lengths in both categories, overall unknown words were slightly longer (mean number of letters = 6.8, *SD* = 1.6) than known words (mean number of letters = 5.1, *SD* = 1.5). High-quality, digital auditory recordings of each word were made using Audacity and edited using Computerized Speech Lab Model 4150 (KayPENTAX). The auditory tokens ranged from 500-1200 ms in length. High-resolution color digital photographs were selected to represent each word. Pre-testing with a separate group of normal adult participants (n = 3) demonstrated that these images accurately represented the corresponding concepts.

2.2. *Task Procedure*

Participants came in for two sessions on two separate days. One session consisted of a visual world task, during which EM and pupillometry data were recorded. The alternative session consisted of a picture-word congruity task, during which ERP data were collected. At the end of the second session, normal adult participants also performed a word familiarity task.

2.2.1. Visual world task

In the visual world task (presented in E-Prime version 2.0.8.74), participants were presented with four pictures, one in each corner of the computer screen, followed 20 ms later by the presentation of an auditory word. Normal adult participants were asked to indicate, using the computer mouse, which picture matched the spoken word. Some LFA participants were better able to maintain attention to the stimuli if given an explicit task; these participants were asked to indicate, using the computer mouse, which picture matched the spoken word. All other LFA participants were asked to sit quietly without moving and look at the pictures.

Pictures representing known words were presented with other pictures representing known words, and pictures representing unknown words with other pictures representing unknown words, so that participants could not eliminate foils in the unknown condition based on familiarity. Each trial began with a fixation cross in the center of the screen (presented for 1000 ms) to ensure that participants' eyes began equidistant from the pictures. The pictures remained on the screen until one was selected with a mouse click or for a maximum of 5000 ms after presentation of the auditory stimulus. The experimental session consisted of 160 trials, one per

experimental item, presented in 8 blocks of 20 trials each. In each block, half of the stimuli were known targets and half were unknown; these were pseudorandomized within blocks. During the visual world task, EM and PD data were collected throughout using an ASL Model 504 eye-tracking system.

2.2.2. Picture-word congruity task

In the picture-word congruency paradigm (also presented in E-Prime), participants were presented with a picture followed 700 ms later by a spoken word. Each known and unknown word and picture was presented twice: once in an incongruent context, in which the word and picture did not match, and once in a congruent context, in which the word and picture matched, yielding 320 trials total. In the incongruent condition, although the picture and the spoken word did not match, they were always drawn from the same knowledge condition (known or unknown). Pairings in the incongruent condition were created such that the picture and the spoken word never shared an initial phoneme. A red fixation point (presented for 1000 ms) began each trial to ensure that participants would be looking at the stimulus when it appeared on the screen. The picture remained on the screen until 1000 ms after the offset of the auditory token, during which time responses were prohibited.

A response screen (a green fixation point) was then presented until participants made a response or for a maximum of 5000 ms. During this time, normal adult participants were asked to indicate, via a button press, whether the word matched the picture. Some LFA participants were better able to maintain attention to the stimuli if given an explicit task; these participants were instructed to press a button to indicate whether the word and picture matched or to press the button after every word presentation. All other LFAs were instructed to sit quietly and watch the pictures.

To minimize artifacts in the EEG signal, participants were instructed to keep their eyes fixated on the center of the screen, to move as little as possible, and to refrain from blinking during the presentation of the picture and the auditory token. High-density ERPs were recorded during the congruency task at 250 Hz using a 256-channel Hydrocel Geodesic Sensor Net and NetStation version 4.3. Impedences were kept under 50kΩ whenever possible.

2.2.3. Word familiarity task

Following the completion of both the forced-choice and the congruity task, at the end of the second session, participants completed a word familiarity post-test (also presented in E-Prime) in which each of the 160 auditory tokens were presented. Using a button press, participants rated their familiarity with all of the words used in the experiment on a scale from 1 (not very familiar at all) to 9 (very familiar), with the option of 0 for words with which they had no familiarity (words they had never heard before their participation in the experiment).

2.3. *Data Analysis*

Eye movement fixation data from the visual world task were analyzed using ASL Results (Applied Science Laboratories, 2009). For each trial, the presentation slide was divided into five

regions of interest, consisting of the fixation cross in the center and the four picture stimuli. Trials in which less than half of the trial was detected by the eye-tracker were removed before analysis or modeling. A fixation was defined as a period of time during which eye gaze remained at a specific location. Fixation onsets were defined by a stable gaze duration for 100 ms or more and a visual angle variation of 1 degree or less. Fixation offsets were defined by three or more sequential samples that deviated from the fixation start location by 1 or more degrees of visual angle. Dwell time was defined as the duration of time spent looking at the named picture, with or without fixation.

Pupil diameters were measured horizontally and recorded in pixels, then converted to millimeters. Small blinks were replaced by linear interpolation. Trials containing 20 or more missing data points in a row (340 ms or more) due to lack of fixations were removed before analysis or modeling. For each trial, a ‘baseline’ pupil diameter, averaged over the 200 ms preceding the stimulus onset, was subtracted from each measurement of the task-evoked pupil diameter.

ERP data were pre-processed using EEGlab version 10.2.2 (Delorme & Makeig, 2004) and Matlab version 8.1 (MathWorks, Inc.). The data were filtered using a 0.1-30Hz bandpass filter and re-referenced to the Cz electrode using an average reference transform. ERPs were time-locked to the onset of the auditory word, and extended from 800 ms before to 1000 ms after the auditory stimulus. Correction for eye movement or motion artifacts was performed using independent component analysis (ICA; Jung et al., 2000). Following ICA decomposition, a joint probability algorithm was used to automatically remove any further bad trials containing eye movements, blinks and other sources of noise.

3. Model training with normal adults

3.1. *Participants*

Participants were 23 adults, all right-handed native English speakers between 19-61 years of age (mean age = 35 years, *SD* = 14; 16 male, 7 female). They all reported normal or corrected-to-normal vision and hearing, and no history of cognitive, learning, or neurological impairments. Participants were recruited from the Johns Hopkins University and Baltimore community. The experimental procedures were approved by the Johns Hopkins School of Medicine Institutional Review Board. All subjects gave written informed consent before participation in the experiment. All received monetary compensation for participating.

3.2. *Modeling procedure*

Modeling was performed using linear mixed effects modeling, implemented using the lme4 package version 1.1-7 (Bates, Maechler, Bolker, & Walker, 2014) with R version 3.2.0 (R Core Team, 2015). The parameter estimation was performed with residual maximum likelihood (REML).

All variables were normalized before modeling. As the intent of this work is to provide a within-subjects estimate of word knowledge based on the implicit measures, normalization was

performed within subjects. Normalizing within subjects has the effect of accounting for variability among participants. For example, consider a hypothetical example in which one participant has relatively large N400 effect magnitudes ranging between 1-5uV, whereas another participant has smaller effects ranging between 0.1-1uV. In this second participant, if a specific word had an N400 effect of 1uV, this would be a large effect compared to the other trials for this participant; however, this would be a small effect compared to other participants. Normalizing within each subject instead ensures that the strength of the effect is considered within the range that is typical for each participant. This procedure can be especially helpful when dealing with data from clinical populations. For example, the implicit measures for an LFA participant might distinguish between known and unknown vocabulary, but the magnitude of effects may be overall smaller than those of a normal adult participant. In such a case, normalization would help to emphasize the differences in the implicit measures between known and unknown vocabulary.

Normalization was performed using the following algorithm:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

where $x = (x_1 \ldots x_n)$ for each individual variable and participant and z_i is the i^{th} normalized data point. These normalized variables were used for both fixed and random effects.

3.2.1. Dependent variable

The dependent variable was the subjective knowledge ratings. After performing both the visual world and the picture-word congruency paradigms, participants rated their knowledge of each word they had encountered on a 10-point scale from 0 (completely unknown) to 9 (completely known). Because a full 10-point rating scale would make accurate model prediction more difficult, the original subjective knowledge ratings ranging from 0-9 were re-scaled to five groups: 1 (ratings of 0 or 1), 2 (ratings of 2 or 3), 3 (ratings of 4 or 5), 4 (ratings of 6 or 7), and 5 (ratings of 8 or 9).

3.2.2. Random effects

The variable of *word* was included as a random effect, as each subject saw the same set of 160 words. Because the ultimate goal of this modeling work is to predict data for new participants, *subject* was not included as a random effect. Although including *subject* as a random effect would account for some of the individual variability among participants, and might lead to a better fit when training the model, only the fixed effects parameters are used when predicting to a new dataset. A regression model with by-subject random intercept and by-subject random slopes cannot generalize to unseen subjects, as no by-subject intercepts or slopes will have been estimated for these subjects. For these reasons, and given that the intention of this work was prediction, *subject* was not included as a random effect.

To account for the possibility that the slopes of some variables may differ for each word, each variable selected for inclusion in the model was tested in a model by itself with only varying intercepts for *word* (the “null” model) and with varying intercepts for *word* and varying slopes for the variable (the “test” model). The null and test models were compared using a chi-squared test; if the test was statistically significant, this indicated that the inclusion of random slope was

needed for that variable. Although the maximal random effects structure would include varying slopes for all variables, this model did not converge; this procedure was used as a way of simplifying the random effects structure in a principled way (Barr, Levy, Scheepers, & Tily, 2013).

Six variables showed significantly better model fits when the slopes were allowed to vary by word: *N400 effect*, *percent fixation duration*, *mean fixation duration, percent dwell on stimulus, number of fixations,* and *Last*. Varying slopes for these variables were included as random effects; however, the maximal model with all six random slopes did not converge. To simplify the random effects structure, we ran six models with only five variables each (i.e. leaving only one variable out each time) and compared models using a criterion-based method by evaluating the Akaike Information Criterion (AIC) for each model. AIC provides an estimate of the goodness-of-fit while penalizing for added complexity. When tested on the same dataset, smaller AIC values represent a relatively better fit. The model with the smallest AIC that had five random slopes and successfully converged was chosen as the final model.

The random effects structure of the final model included varying slopes for the effects of *N400 effect, percent fixation duration time, average fixation duration time, percent dwell,* and *number of fixations* by word. Random effects were modeled using an unstructured covariance matrix.

3.2.3. Independent variables/fixed effects

The independent variables were 13 measures taken from the EM, PD, and ERP data (see Appendix 1 for intercorrelation[1] matrix). These measures were entered as main effect terms into the model. Because the aim of this work was prediction rather than interpretation, all of these variables were included as fixed effects; we did not perform a variable selection step to determine which variables would be included in the model. Including all of the variables allows for model prediction to use as much data as possible, which might be useful when extending to new subjects.

3.2.3.1. EM measures

A number of independent variables were taken from the eye-movement measures.

- *Total number of fixations* was the total number of fixations made during the trial.
- *Mean fixation duration on the stimulus* was the average time (in ms) throughout the trial spent fixating on the target stimulus.
- *First fixation duration* was the time (in ms) of the first fixation on the target stimulus.
- *First dwell* was the cumulative time (in ms), including all fixations and saccades, of the first entry into the quadrant of the target picture before leaving that quadrant.

[1] We did not account for multicollinearity in the data. Collinearity can cause problems when using stepwise model selection procedures or when attempting to interpret the significance of a specific predictor in a model. However, as we are using mixed effects modeling for prediction purposes rather than interpretation, and because we use criterion-based model selection procedures, which are not affected by multicollinearity, this is not an issue in the current endeavor (Cohen, Cohen, West, & Aiken, 2003; Dormann et al., 2013).

- *Latency to first fixation* was the length of time (in ms) taken to first fixate on the target picture.[2]
- *Percent fixation duration on the stimulus* was calculated as the percentage of the entire trial length spent fixating on the target picture.
- *Percent dwell on stimulus* was the percentage of the total trial spent dwelling on the target stimulus.
- *Percentage of trials first fixated ("First")* was whether the target stimulus was the first picture fixated at the start of the trial.
- *Percentage of trials last fixated ("Last")* was whether the target stimulus was the last picture fixated before the response.

3.2.3.2.PD measures

The pupillary dilation measures were *PD: maximum change*, calculated as the largest absolute change in pupil size from baseline; *PD: mean change*, calculated as the average change in pupil size over the entire trial; and *PD: percent change*, calculated as the absolute maximum percent change from baseline over the entire trial.

3.2.3.3.ERP measures

The ERP measure included the *N400 effect*, defined as the magnitude of the N400 effect (in µV) for each word. For each participant and for each presentation of a single word, a difference wave was calculated by subtracting congruent amplitudes from incongruent amplitudes at electrode Pz. The peak negative amplitude was then identified within a window from 200-800 ms after word presentation. A 100 ms window around this peak was taken as the individual N400 window, and the average amplitude of the difference wave was calculated for that window. This average difference wave amplitude was used as the N400 effect.

3.3.*Model training results*

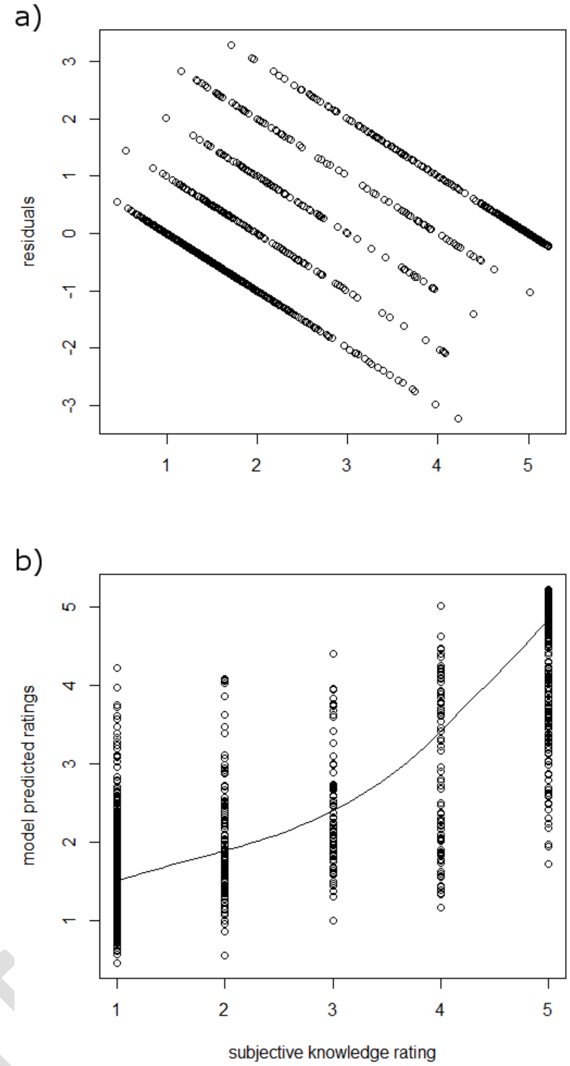
A linear mixed-effects model was fit to the full dataset of normal adults (n=23). To summarize, the random effects in the model included varying intercepts for *word*, and varying slopes for the effects of *number of fixations, mean fixation duration, percent fixation duration*, *percent dwell on stimulus,* and *N400 effect* by word; the fixed effects in the model included all 13 variables. The results of the model are presented in Table 1. Plots of the fitted versus residual values and of the observed knowledge classifications vs. predicted probabilities are presented in Figure 1.

[2] Note that Ledoux et al. (2015) also report *Latency to first refixation*, defined as the amount of time that passed before the target picture was refixated. However, this variable was not included in the model because of a high number of trials containing missing data (if a re-fixation did not occur), which would have caused problems with the modeling (see section 5.3 in Discussion).

Table 1: Random and fixed effects of final model.

Random effects			
Groups	*Name*	*Variance*	*Std. Dev.*
word	(Intercept)	2.53	1.59
	number of fixations	0.53	0.73
	mean fixation duration	0.35	0.59
	percent fixation duration	0.52	0.72
	percent dwell	0.03	0.19
	N400 effect	0.31	0.56
Residual		0.51	0.71
Fixed effects			
	Estimate	*Std. Error*	*t-value*
(Intercept)	3.24	0.17	19.54
number of fixations	-0.60	0.12	-4.92
mean fixation duration	-0.26	0.16	-1.65
first fixation duration	-0.07	0.13	-0.56
first dwell	-0.32	0.13	-2.52
latency to first fixation	0.23	0.12	1.93
percent fixation duration	0.77	0.14	5.44
percent dwell	0.28	0.18	1.57
first	-0.13	0.05	-2.61
last	0.16	0.05	3.15
PD: max change	0.66	0.31	2.12
PD: mean change	0.23	0.10	2.17
PD: percent change	-0.86	0.31	-2.73
N400 effect	0.04	0.10	0.46

Figure 1: Training dataset: a) Plot of the model residuals vs. fitted knowledge classifications; b) Plot of subjective knowledge ratings against model predicted probabilities, with a locally-weighted loess regression line.



3.3.1. Cross-validation

Model validation was performed using leave-one-out cross-validation. This technique allows an estimate of how well this model would predict new data. For each subject, a model was fit using the rest of the dataset (n=22), and predicted values were generated for the subject's data. At each iteration we calculated the root mean squared error (RMSE) as the square root of the averaged squared differences between observed and predicted[3] values. The averaged RMSE over all subjects provided a measure of overall model fit.

The results of cross-validation are shown in Table 2. Although the error rates vary between participants, the overall RMSE for the full model was 0.76. RMSE values are given in the same unit as that of the dependent variable. This means that on average, predicted values were within one rating point of actual ratings.

[3] Because a linear regression model was used, the predicted values were not even integers. For example, given an observed knowledge rating of 3, the model's predicted rating may be 3.57. The original predicted values were used when calculating the error rate in order to better capture the model's error rate. To convert these predicted values to knowledge categories, as in Table 2, the predicted ratings were rounded to the nearest integer.

Table 2: Number of trials included in the model for each subject; percentage of trials in each of the five knowledge rating categories for subjective ratings and model predicted ratings; and the model error rate, for each participant. Note that objective and subjective percentages were calculated only for the trials that were included in the modeling (see section 5.3 of the Discussion regarding missing data).

participant number	# trials included in model (out of 160)	% trials in each knowledge category: subjective					% trials in each knowledge category: predicted					root mean squared error (RMSE)
		1	2	3	4	5	1	2	3	4	5	
1	108	38	0	2	1	59	20	18	2	1	59	0.54
2	143	27	9	4	6	55	18	20	5	6	51	0.66
3	124	48	0	2	1	49	27	11	7	5	49	0.79
4	83	33	7	2	4	54	14	22	11	5	48	0.74
5	149	30	7	3	3	58	15	21	5	7	52	0.91
6	84	30	0	2	7	61	20	13	4	6	57	0.85
7	98	35	3	3	5	54	14	21	7	5	52	0.89
8	115	22	6	6	5	61	18	23	9	6	43	1.07
9	62	32	0	5	0	63	24	10	3	2	61	0.57
10	146	30	6	5	4	54	14	25	3	3	54	0.73
11	30	40	0	0	0	60	27	10	3	3	57	0.43
12	95	33	1	0	3	63	22	18	4	6	49	1.04
13	88	22	3	8	3	64	16	17	8	7	52	0.92
14	79	25	3	0	4	68	10	19	5	6	59	0.63
15	90	36	3	2	2	57	23	11	7	3	56	0.75
16	77	5	38	6	0	51	22	19	4	4	51	0.61
17	139	27	11	2	2	58	14	20	9	6	52	0.78
18	102	41	4	2	0	53	29	15	4	0	52	0.47
19	111	25	2	7	11	55	27	18	1	3	51	1.13
20	107	30	7	5	4	55	18	22	4	1	55	0.69
21	45	47	2	0	0	51	27	11	13	2	47	0.72
22	118	13	14	6	9	58	16	20	4	8	52	0.90
23	57	30	4	2	0	65	9	18	5	5	63	0.64
Average	**98**	**30**	**6**	**3**	**3**	**58**	**19**	**18**	**6**	**4**	**53**	**0.76**

3.3.2.Word frequency effects

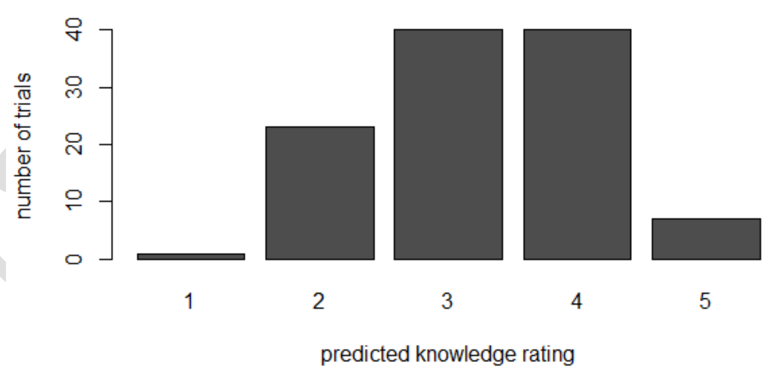
One criticism of this modeling procedure might be that we did not account for word frequency effects. All of the words used in the two tasks were either highly frequent or highly infrequent; word frequency was in fact used to split the data into the original objective categories of “known” and “unknown”. One possible explanation for these results may be that the regression

model is simply capturing word frequency effects rather genuine subjective vocabulary knowledge.

One way to address this question is to examine trials in which the word was low-frequency (i.e. objectively-labeled "unknown" words) but the participant rated it as "known" (i.e. gave it a rating of 5). Over all normal adult participants, there were 111 such trials. If the model is only capturing word frequency effects, the model's predicted ratings for these words should be close to or equal to 1 ("unknown"). However, if the model is capturing subjective word knowledge based on the patterns of the implicit data, the model's predicted ratings should more closely align with the subjective ratings of 5.

Examining the model's predicted ratings for this subset of words that were low-frequency but had a subjective rating of 5 showed that the predicted ratings fell more towards the higher knowledge ratings (see Figure 2). Of the 111 trials, 7 trials (6%) were predicted as being in category 5; 40 trials (36%) were predicted as being in categories 4 and 3 each; 23 trials (21%) were predicted as being in category 2; and only 1 trial (1%) was predicted as being in category 1. Therefore although there was some error, overall the model was fairly successful at predicting the subjective knowledge ratings even when words were low-frequency. This suggests that the model is not just capturing word frequency effects, but is using the implicit measures to predict subjective knowledge.

Figure 2: Predicted knowledge ratings for words that were low-frequency (objectively-rated "unknown" trials) but had a subjective knowledge rating of 5 ("known"). The fact that most trials did not have a predicted knowledge rating of 1 suggests that the model is capturing subjective knowledge rather than just word frequency effects.



4. Extension to a low-functioning population

The ultimate aim of this work is to be able to predict receptive vocabulary knowledge in a population in which assessment of such knowledge is difficult or impossible to obtain. To demonstrate the feasibility of using regression models to predict vocabulary knowledge in this

population, we extended the model to a group of low-functioning individuals with autism (LFAs).

4.1.*Participants*

Participants were 5 low-functioning individuals with autism. Their mean age was 32 years (SD = 15; range 18-48); all males; 4 Caucasian, 2 Asian. All had normal or corrected-to-normal vision and hearing. Participants were recruited from the Johns Hopkins University and Baltimore community. The experimental procedures were approved by the Johns Hopkins School of Medicine Institutional Review Board. Written informed consent was obtained from each participant and their legal guardian before participation in the experiment. All received monetary compensation for participating.

Criteria for identifying these participants as LFAs were based on the severity of core features of autism as stated in DSM-5; the severity of environmental support and supervision needed; and (if applicable) the total score from the Autism Diagnostic Observation Schedule (ADOS). Although intelligence, receptive language, and self-injurious or aggressive behaviors were assessed and documented, they were noted as possible associated features of autism rather than core features of autism. Although these associated features were included in obtaining an overall picture of each participant, they were not included in identifying these individuals as low-functioning. While all participants required 24-hour support staff and were classified as LFAs, they varied greatly in the severity of their symptoms and the range of their intelligence and verbal abilities (see Table 3). For these reasons, in this paper we reject intellectual and verbal ability as characteristic of low-functioning status. Rather, we define low-functioning autism according to DSM-5 Level 3 (Severe Level of Autism), which marks severe deficits in social communication and restricted and repetitive behaviors requiring substantial support throughout the individual's daily life. All participants exhibited restricted and repetitive behaviors and severe deficits in verbal and/or nonverbal social communication skills that significantly affected their level of daily functioning. Direct 24-hour support staff and/or parental supervision, with a focus on activities of daily living and functional communication, was required for each participant. All participants were enrolled in adult or educational programs specific to assisting individuals with autism, including five participants from the Linwood Center in Baltimore, a program that provides residential and educational services for individuals with autism. The Linwood Center provides child and adult services for individuals living with autism. The Adult Services Program includes Supported Employment, the Day Habilitation Program, and Residential Services. The Linwood Center is an approved IRB Research site and has been instrumental in research involving individuals on the autism spectrum.

All participants had a current diagnosis of autism as confirmed by record review. To verify the diagnosis, we administered the Autistic Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) and Autism Diagnostic Observation Schedule (First Edition (ADOS-1) or Second Edition (ADOS-2), depending on the current version of the assessment at the time of testing; Lord et al., 2000). These assessments were administered by members of the research team who had completed the official ADOS training and who have extensive experience working with individuals on the autism spectrum in both research and educational settings. The Kaufman Brief Intelligence Test, Second Edition (K-BIT-2; Kaufman & Kaufman, 2004), was

administered to assess verbal and non-verbal intelligence. The Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007), was administered to assess receptive vocabulary.

Table 3 shows the results of these neuropsychological assessments for all participants. Participants unable to complete the assessments were either non-compliant with the testing protocol or incapable of making reliable responses. The ADI-R could not be obtained for three of the individuals because they were adults in assisted living programs. For participants for whom the ADI-R could be completed, all assessments confirmed the diagnosis of autism.

For two of the five participants there was no appropriate module of the ADOS[4]. (Currently no modules address nonverbal adolescents or adults, although Adapted Modules 1 and 2, for non-verbal individuals 18 years of age and older, are being developed.) Each of these six participants either did not meet criteria for expressive language skills for a specific module (regardless of chronological age) or the ADOS module that met criteria for expressive language skills was developmentally inappropriate for the participant's chronological age. For these participants, the researchers performed "adapted" modules by interacting with the participants and identifying the specific behaviors measured by the ADOS. These adapted scores are noted in Table 3, but we caution that they cannot be considered "official" ADOS scores.

Two of the five LFA participants in our sample were non-verbal and were unable to provide accurate behavioral assessments of their cognitive abilities. These individuals represent the population we would most like to target with this work. However, even though the other participants in this sample were more verbal, additional testing issues such as lack of motivation, haphazard patterns of behavioral responses, and difficulty using equipment like a response or mouse could contributed to the difficulty in obtaining accurate behavioral assessments of cognitive abilities. Participants of varying levels of verbal and cognitive abilities were included in this sample of LFAs to illustrate that the modeling procedure we describe is not limited to non-verbal individuals but might be useful for low-functioning populations in general.

[4] There are five possible modules of the ADOS: The Toddler Module is used with children between 12-30 months who do not consistently use speech; Module 1 is used with children 31 months and older with little-to-no speech; Module 2 is used with children who have some speech but who are not verbally fluent; Module 3 is used with children and adolescents who are verbally fluent; and Module 4 is used with adolescents and adults who are verbally fluent

Table 3: Participant demographics, including autism diagnostic test scores (ADOS, ADI-R), intelligence scores (K-BIT), and vocabulary scores (PPVT). N/A indicates that test was not appropriate for the individual's level of functioning or could not be completed. Note that symptom severity scores for the ADOS are not given for Module 4. *The ADOS-1 does not give symptom severity so total scores were compared with the ADOS-2 algorithm.

Participant number	**ADOS**					**ADI-R**	**K-BIT**	
	ADOS version	**Module**	**Total**	**Classification**	**Symptom Severity**		**verbal**	**non-verbal**
1	1	1 (adapted)	20	autism	high*	completed	N/A	
2	N/A					completed	N/A	
3	2	4 (adapted)	22	autism	--	N/A	45	79
4	2	4	20	autism	--	N/A	40	60
5	2	4	19	autism	--	N/A	93	131

4.2.*Model prediction procedure*

All LFA participants performed the visual world task and the picture-word congruity tasks described above (see section 2.2) while EM, PD, and ERP data were recorded. Data for each of the 13 variables described above were first normalized for each measure and subject (see section 3.2). These data were then entered into the previously-built model to generate predicted knowledge ratings.

4.3.*Model prediction results*

Table 4 shows the distribution of trials for each participant that were objectively labeled as "known" and "unknown" and predicted for each category. Examination of the predicted knowledge ratings for the LFA participants showed that all objectively-labeled "known" words were given predicted ratings of 4 or 5. This suggests that all LFA participants were familiar with the high-frequency words used in the current paradigms. The predicted knowledge ratings for the low-frequency "unknown" words showed a slightly wider range of values; the majority of trials had predicted ratings of 1 or 2, suggesting that participants were indeed unfamiliar with the unknown words. As can be seen in Table 4, some LFA participants had only a few trials that had enough good data on all measures to be able to predict a knowledge rating. We will discuss this in more detail in section 5.3 of the Discussion.

Figure 3: Results of model predictions for the LFA participants. For each participant, plots are shown of the predicted knowledge ratings for each of the objectively-rated "known" and "unknown" word trials. Each individual trial is shown as a grey dot. Boxplots show the median value in each objective category and the quartiles of the distribution.

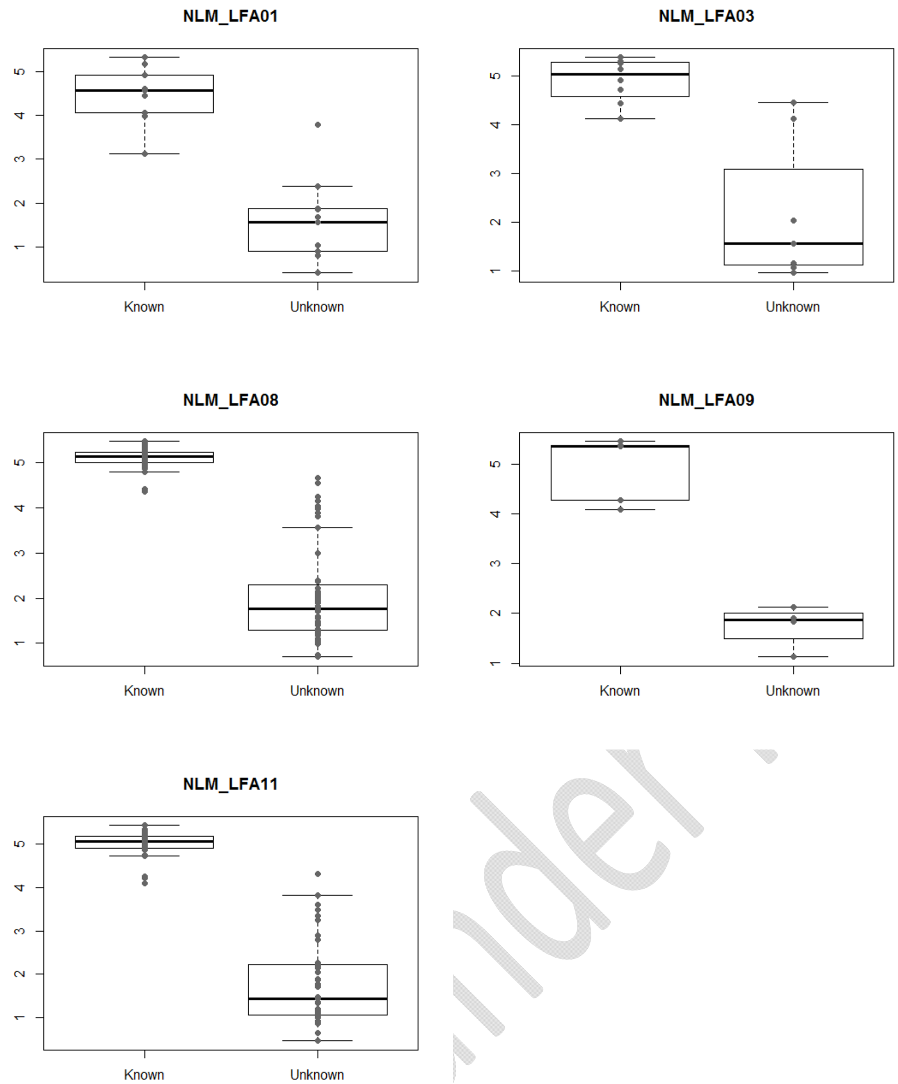


Table 4: Number of trials included in the model for each subject; percentage of trials objectively labeled as “known” and “unknown”; and percentage of trials in each of the five knowledge rating categories for model predicted ratings.

participant	# predictable trials (out of 160)	% trials in each objective knowledge category		% trials in each knowledge category: predicted				
		“known”	**“unknown”**	**1**	**2**	**3**	**4**	**5**
1	20	45	55	25	25	5	20	25
2	16	50	50	19	19	0	25	38
3	102	54	46	18	19	1	10	53
4	9	56	44	11	33	0	22	33
5	73	52	48	26	11	7	11	45

5. Discussion

The current study aimed to demonstrate that implicit measures of EM, PD and ERPs can be used to estimate latent vocabulary knowledge through the use of linear mixed effects regression modeling. In two sessions, participants performed a visual world task while EM and PD data were collected, and a picture-word congruity task while EEG data were collected. We first fit a regression model to a dataset of normal adult participants to capture the relationship between three implicit measures of receptive vocabulary and subjective word knowledge ratings provided by these participants. After training the model on the normal adult data, we then used the model to predict knowledge ratings for a population of low-functioning individuals with autism using only their implicit measures.

5.1.*Model training with normal adults*

A linear mixed effects model was first fit to a dataset of 23 normal adult participants. The dependent variable was subjective knowledge ratings provided by each participant, and the independent variables were 13 measures taken from the EM, PD, and ERP data. Random effects of word and random slopes for five variables were also included.

Leave-one-out cross-validation demonstrated that the model was very successful at predicting receptive knowledge from implicit EM, PD, and ERP measures. The overall root mean squared error was 0.76; even though error rates varied between participants, in no cases did the error rate exceed 1.2. This indicates that the predicted knowledge ratings fell within about one rating point of actual subjective ratings in most cases. This model training and validation in normal adults therefore demonstrates that the regression model can accurately capture subjective vocabulary knowledge.

Importantly, the model does not seem to be reflecting mere word frequency effects. We examined a subset of trials in the normal adult data on which low-frequency “unknown” words were given a subjective knowledge rating of 5, suggesting that the participants were familiar with these low-frequency words. If the model were capturing word frequency effects only, the predicted ratings for these words would be expected to fall around 1 (“unknown” categories). However, for these trials, the predicted knowledge ratings fell more towards the higher end of the knowledge ratings, suggesting that the model was accurately capturing subjective word knowledge rather than word frequency effects.

5.2.*Extension to LFAs*

The second aim of this study was to use the regression model that was trained on the normal adult data to predict vocabulary knowledge from the implicit measures in the absence of overt behavioral responses. To do so, we entered the EM, PD, and ERP data from a group of LFAs into the regression model and generated predicted knowledge ratings. Overall, the predicted ratings showed that the high-frequency words that we had expected to be “known” to most participants were also familiar to this group of LFAs. For unknown words, the majority of predicted ratings fell in categories 1 and 2, although there was slightly more spread, with some

predicted ratings of 3 or 4. This may reflect a "ceiling effect" for known words, such that all are predicted to fall within knowledge categories of 4 or 5, whereas there is more variability for unknown words. On the whole, this work demonstrates that regression modeling can be used to predict latent vocabulary knowledge in individuals who may be unable to provide accurate estimates of their knowledge through an overt behavioral response.

Of course, we cannot be certain of the accuracy of these predictions. As the model performed well with estimating receptive vocabulary in the normal adult population, we have reason to believe that it is also fairly accurate at predicting vocabulary knowledge in the LFA participants. However, as the LFA participants did not – and in some cases, could not – provide explicit reports of their knowledge, we cannot be sure that these predictions reflect their true knowledge. Yet an estimate of receptive vocabulary abilities based on cognitive measures, even if slightly inaccurate, is better than nothing. In this way, moving towards a more quantitative estimate language ability based on patterns of cognitive functioning provides a basis for assessing vocabulary knowledge in patient populations.

These estimates of vocabulary ability could also be useful for assessing the results of language interventions in clinical populations. For example, a vocabulary training program might collect these implicit measures before and after intervention, then assess how predicted knowledge ratings have changed after intervention. Words that are given a predicted knowledge rating of 5 after intervention might be classified as "learned" and could be moved out of the training pool, whereas words that maintain a predicted knowledge rating of 2 or 3 even after intervention might be classified as "still needs work" and could undergo more training. In this way, this technique of predicting knowledge ratings for specific words could be used to tailor intervention programs for a specific individual and word set.

5.3. *The issue of missing data*

One important limitation that stems from this more complex mixed model is that if any word had missing data for any single independent variable, the model was unable to generate a predicted knowledge value for that word. This resulted in a high amount of data loss, as can be seen in Table 2 and Table 4. The rate of data loss is higher in the LFA participants, with one participant having only 9 trials on which good data for all measures led to the ability to generate predicted knowledge ratings. Techniques of multiple imputation do exist for replacing missing data in regression models; however, this procedure becomes much more complex with mixed effects models and with prediction to new datasets. This is an active area of research in mixed regression modeling, and we anticipate that future developments will allow for recovery of missing data and more accurate model fitting.

Both EEG and eye-tracking methodologies are extremely sensitive to movement, which led to the high degree of data loss in some participants. This stresses the importance of collecting clean data from the outset: any and all attempts to maximize the number of good trials should be made during data collection. In the case of eye-tracking and ERP measures, this mainly consists of minimizing movement, so techniques of ensuring participant comfort and engagement in the task are important. Offline data cleaning procedures can and should also be utilized to ensure the cleanest data possible while maximizing data retention. These procedures are especially

important for patient populations, which might suffer higher rates of data loss due to the inherent difficulties in testing these populations. In such cases, some modifications to data acquisition and cleaning procedures may be needed (e.g. see Kylliäinen, Jones, Gomot, Warreyn, & Falck-Ytter, 2014).

The issue of missing data demonstrates an important point for future research that intends to use this modeling procedure to provide an estimate of receptive vocabulary knowledge for a specific word or set of words. Given the high potential for missing data due to messy trials, it will be important to collect multiple datapoints for each word of interest. Increasing the number of presentations of the word(s) of interest will ensure that good data is collected for at least one exposure, so that accurate knowledge predictions can be made.

6. Conclusions

In sum, the current work demonstrates that mixed effects models can be used to predict latent receptive vocabulary knowledge from implicit assessment techniques of eye-tracking, pupillary dilation, and event-related potentials even in the absence of behavioral responses. The ability to estimate vocabulary knowledge is of immense importance for populations in whom assessment of receptive capacity is difficult, such as non-verbal individuals with autism. This methodology is by no means limited to the specific research areas or technologies employed in the current study: a similar approach of fitting a regression model to a normal population and using it for prediction to a clinical population could be used to investigate virtually any aspect of comprehension or cognition, such as memory, reasoning, or consciousness. These paradigms and procedures could also easily be extended to alternative technologies such as functional magnetic resonance imaging (fMRI). This work offers a proof-of-concept demonstration of the use of regression modeling to predict cognitive abilities from implicit measures, which holds great potential to improve the assessment of cognition in patient populations.

REFERENCES

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from http://cran.r-project.org/package=lme4

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (2nd ed., pp. 142–162). New York, NY: Cambridge University Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Connolly, J. F., & D'Arcy, R. C. (2000). Innovations in neuropsychological assessment using event-related brain potentials. *International Journal of Psychophysiology*, *37*, 31–47.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*, 27–46.

Dunn, L., & Dunn, D. (2007). Peabody Picture Vocabulary Tests (4th edition). *American Guidance Service*. Circle Pines, MN: American Guidance Service.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465–1477.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461.

Hyönä, J., Tommola, J., & Alaja, A. M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal Of Experimental Psychology*, *48A*(3), 598–612.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Kaufman, A., & Kaufman, N. (2004). Kaufman Brief Intelligence Test (2nd edition). Circle Pines, MN: American Guidance Service.

Kuipers, J.-R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *5*(61), 1–5.

Kuipers, J.-R., & Thierry, G. (2013). ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex*, *49*(10), 2853–2860.

Kutas, M., & Hillyard, S. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205.

Kylliäinen, A., Jones, E. J. H., Gomot, M., Warreyn, P., & Falck-Ytter, T. (2014). Practical Guidelines for Studying Young Children With Autism Spectrum Disorder in Psychophysiological Experiments. *Review Journal of Autism and Developmental Disorders*.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933.

Ledoux, K., Coderre, E. L., Bosley, L., Buz, E., Gangopadhyay, I., & Gordon, B. (2015). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavior Research Methods*. doi:10.3758/s13428-015-0571-6

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685.

Odekar, A., Hallowell, B., Kruse, H., Moates, D., & Lee, C.-Y. (2009). Validity of eye movement methods and indices for capturing semantic (associative) priming effects. *Journal of Speech, Language, and Hearing Research*, *52*, 31–48.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rugg, M., & Coles, M. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition.* Oxford University Press.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, *17*, 133–145.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493.

Tanenhaus, M., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Turner, L. M., Stone, W. L., Pozdol, S. L., & Coonrod, E. E. (2006). Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism*, *10*(3), 243–265.

ACKNOWLEDGEMENTS

The authors would like to thank Esteban Buz, Ishanti Gangopadhyay, and Laura Bosley for their help with data collection, as well as Rasika Jayatillake, Zachary Fisher, and Bonnie Nozari for their help with statistical modeling.

This research was supported by a grant from the Nancy Lurie Marks Foundation; the Department of Defense Autism Research Program Grant # AR093137; The Therapeutic Cognitive Neuroscience Fund; The Therapeutic Cognitive Neuroscience endowment; and the Benjamin and Adith Miller Family Endowment on Aging, Alzheimer's, and Autism Research.

APPENDICES

Appendix 1: Intercorrelation matrix (using Spearman correlations) of the fixed effects for the normal adult model training dataset (n=23). All variables were normalized within each subject before running correlations. *First* and l*ast* were originally coded as Y or N, so they were re-coded as binary variables 0 (N) or 1 (Y) for the purposes of correlation analyses.

	number of fixations	mean fixation duration	firstpass fixation duration	first dwell	latency to first fixation	percent fixation duration	percent dwell	first	last	PD: max change	PD: average change	PD: percent change	N400 effect
number of fixations	1												
mean fixation duration	-0.21	1											
firstpass fixation duration	-0.22	0.67	1										
first dwell	-0.25	0.38	0.43	1									
latency to first fixation	0.20	-0.01	0.05	-0.09	1								
percent fixation duration	-0.57	0.31	0.19	0.47	-0.37	1							
percent dwell	-0.46	0.38	0.23	0.65	-0.43	0.83	1						
first	-0.11	-0.03	-0.13	0.03	-0.58	0.44	0.28	1					
last	-0.50	0.21	0.14	0.22	-0.09	0.58	0.52	0.04	1				
PD: max change	0.17	-0.01	-0.07	-0.12	0.09	-0.18	-0.20	0.02	-0.19	1			
PD: average change	0.16	0.03	-0.03	-0.04	0.01	-0.11	-0.06	0.05	-0.12	0.39	1		
PD: percent change	0.17	0.00	-0.07	-0.12	0.10	-0.17	-0.19	0.03	-0.18	0.97	0.45	1	
N400 effect	0.09	0.04	0.03	-0.04	0.10	-0.07	-0.03	-0.05	-0.02	0.01	-0.02	0.00	1

Appendix 4

Gangopadhyay, I., Ledoux, K., Bosley, L., & Gordon, B. (2012, April). The Use of Implicit Measures to Assess Vocabulary Knowledge in Normal Adults and Normally Developing Children. Poster presented at the 19th Annual Meeting of the Cognitive Neuroscience Society, Chicago, IL.

The Use of Implicit Measures to Assess Receptive Vocabulary Knowledge in Normal Adults and Normally Developing Children

Gangopadhyay I[1], Ledoux K[1], Bosley L[1], Gordon B[1,2]
[1]Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD;
[2]Department of Cognitive Science, The Johns Hopkins University, Baltimore, MD

Introduction

An important question about assessing language comprehension is whether we can use implicit measures to detect evidence of receptive vocabulary knowledge in the absence of explicit behavioral responses. In this study we use event-related potentials (ERPs), pupillary dilation (PD), and eye movements (EMs) as measures of receptive vocabulary knowledge in two groups – normal adults and normally developing children, in whom these implicit measures could be validated by explicit behavioral responses.

Event-related potentials (ERPs): The N400 component of ERP waveforms has been associated with semantic processing, such that words or pictures that are semantically congruent with their proceeding context elicit a smaller-amplitude N400 than words or pictures that are incongruent; this difference has been called the N400 congruency effect (Connolly & D'Arcy, 1999).

Pupillary dilation monitoring (PD): Task-specific changes in pupillary diameter that are time-locked to the onset of events (stimuli or responses) have long been associated with attentional engagement and information processing. Pupillary dilation has been shown to increase with task difficulty in many tasks, and has thus been taken as a measure of resource recruitment (Beatty & Lucero-Wagoner, 2000).

Eye movement monitoring (EM): Eye movements typically reflect current cognitive operations. For example, participants will look at objects in a display as they hear the names of those objects. Studies of normally-developing children have suggested that such eye movements become faster and more precise as children learn the meanings of spoken words (Swingley & Fernald, 2002).

PARTICIPANTS

➢ ***20 normal adults:***
- Right-handed native English speakers
- Normal/corrected-to-normal vision
- 18 years and older

➢ ***16 normally developing children:***
- Right-handed native English speakers
- Normal/corrected-to-normal vision
- 5-17 years of age

➢ ***All participants scored within the normal ranges of the PPVT and KBIT for verbal knowledge.***

METHODS

***Stimuli*:**

➢ 160 word and picture pairs
- 80 "known" (ex. *airplane* and *camera*)
- 80 "unknown" (ex. *agouti* and *cainito*)

***Tasks*:**

➢ ERP congruity task: a picture was presented on the computer screen, along with the auditory presentation of a single word (known or unknown), which matched (congruous) or did not match (incongruous) the visually presented item. Participants were asked to push a button to indicate whether the word and picture matched.

➢ Eye-tracking forced-choice task: participants were asked to select one of the four pictures presented simultaneously on the computer screen after hearing one of the objects named.

➢ Behavioral responses served as comparisons for implicit measures.

Results - Adults

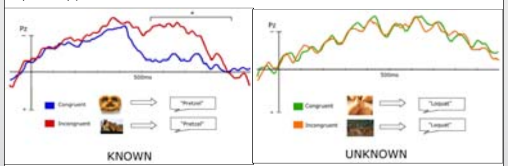
Normal adults:

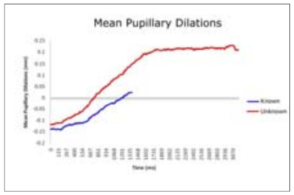
ERP: ANOVAs were performed on mean amplitudes at 50ms intervals with knowledge (known, unknown), congruency (congruent, incongruent) electrode site (frontal, central, parietal) and laterality (left, right) as factors.

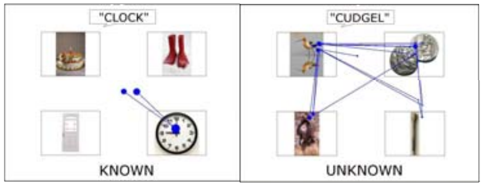
As predicted, a significant N400 congruency effect was observed from 550 to 900ms only for known words, all $F(1,19) > 4.5$, $p < 0.05$. There was no significant laterality interaction in any time interval. However, the effect was significantly larger in posterior locations $p < 0.01$.

PD: PDs from baseline were greater in the unknown condition. The mean peak dilation was markedly greater for unknown words (M=.81 mm) than known (M=.46 mm) $p < 0.01$, indicating that the average change in pupil size was greater for the unknown items.

EM: EMs were faster to pictures for known (760ms) compared to unknown words (1060ms) $p < 0.01$. End-of-trial fixations were on the named picture more frequently for known (94.0%) than unknown words (34.1%) $p < 0.01$.







Results - Children

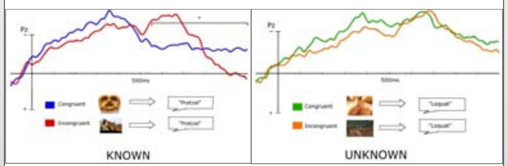
Normal children:

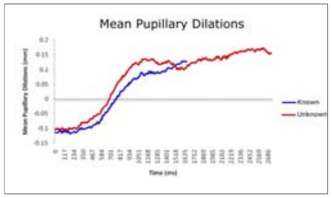
ERP: Similar ANOVAs were performed with 16 children data sets.

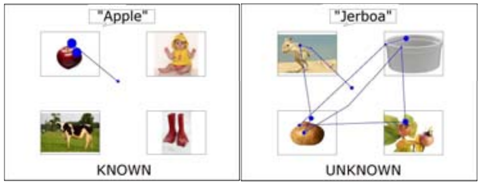
A significant N400 congruency effect was observed from 550 to 1000ms only for known words, all $F(1,15) > 4.7$, $p < 0.05$. There was no significant laterality interaction in any time interval. However, the effect was significantly larger in posterior locations $p < 0.05$.

PD: PDs were significantly greater in the unknown condition. The mean peak dilation was significantly greater for unknown words (M=.71 mm) than known words (M=.44 mm) $p < 0.01$, indicating that the average change in pupil size was greater for the unknown items.

EM: EMs were faster to pictures for known (830ms) compared to unknown words (1060ms) $p = 0.02$. End-of-trial fixations were on the named picture more frequently for known words (94.7%) than unknown words (32.0%) $p < 0.01$.







Conclusions

Adults:
- An N400 congruency effect was observed in the adults, but only for words that were expected to be known.
- The effect had a greater posterior distribution with no significant hemispheric differences.
- No N400 congruency effect was observed for the unknown condition.
- Changes in pupillary dilation were greater for unknown words, relative to known words, suggesting greater attentional engagement.
- Eye movements were faster and more accurate for known words than for unknown words.

Children:
- There was a significant N400 congruency effect only to the known words.
- The effect was more posterior with no laterality differentiation.
- Changes in pupillary dilation were greater for unknown words.
- Eye movements were faster and more accurate for known words.

Although the results were comparable, there were also noticeable differences between the two groups. The children showed a later positivity to the known congruent condition in the ERP, which was absent in the adults. Additionally, the children showed a steady increase in their PDs for known words and the differences between the known and unknown words were smaller, compared to the adults. These data suggest that adults and children might have different cognitive processing for known words.

From the results above, we can conclude that ERPs, PD, and EMs are capable of assessing single word comprehension. Due to its consistency between adults and children, we also predict that eye movements might be the best indicator of receptive word knowledge. And although different processes might be occurring in the two groups, all three techniques are still valid methods for differentiating known from unknown words.

These results also propose an effective way of assessing word comprehension in populations that are minimally verbal or nonverbal. We are currently in the process of testing low-functioning individuals with autism, a population that has been difficult to evaluate due to insufficient responding, poor motivation, and various other behavioral deficits. All three measures (ERPs, PD, and EM) will be useful in assessing language comprehension in such individuals who are unable to make overt behavioral responses.

Bibliography

Connolly, J., & D'Arcy, R. (1999). Innovations in neuropsychological assessment using event-related brain potentials. *International Journal of Psychophysiology, 37*, 31-47.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J.T. Cacioppo, L.G. Tassinary, G.G. Berntson (Eds.), *Handbook of psychophysiology*, 2nd edition (pp. 142-162). New York, NY: Cambridge University Press.

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language, 46*, 39-56.

EQUIPMENT

•EM/PD: Applied Scientific Laboratories 504 Eye-Tracking System
•ERP: Electrical Geodesics Inc. GES 300 EEG System with 256-channel Hydrocel Geodesic Sensor Nets

ACKNOWLEDGEMENTS
This work was supported by The Nancy Lurie Marks Family Foundation and by the Department of Defense grant award W81XWH-10-1-0404.

Appendix 5

Coderre, E., Gordon, B., & Ledoux, K. (2014, April). Neural Connectivity During Semantic Processing of Pictures and Spoken Words in Autism Spectrum Disorders. Poster presented at the 21st Annual Meeting of the Cognitive Neuroscience Society, Boston, Massachusetts.

Neural connectivity during semantic processing of pictures and spoken words in autism spectrum disorders

Emily Coderre[1], Barry Gordon[1,2], & Kerry Ledoux[1]
[1]*Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD*
[2]*Department of Cognitive Science, The Johns Hopkins University, Baltimore, MD*

Introduction

Autism is a pervasive developmental disorder that manifests in a wide variety of cognitive deficits. Complex language is especially impaired in autism spectrum disorders (ASD), particularly higher-level functions like semantic integration and pragmatics (Tager-Flusberg et al. 2005).

In ERP studies of language processing, the N400 indexes semantic integration, showing a reduced negative amplitude for congruent semantic contexts compared to incongruent contexts (Kutas & Federmeier, 2010). Individuals with ASD show reduced or absent N400 effects compared to controls (e.g. McCleery et al., 2010), suggesting impaired semantic integration.

There is also evidence for a general pattern of underconnectivity, both during the resting state and during language processing, in ASD compared to controls (Just et al., 2004). Specifically, underconnectivity between left fronto-parietal networks may contribute to the observed deficits in higher-level language in ASD (Jones et al, 2010).

However, all previous studies investigating connectivity during language processing in ASD have used fMRI. EEG coherence analysis is better suited to capture the dynamic changes in neural connectivity during semantic processing. Only one study has performed spectral analyses of EEG data during a language processing task in ASD (Braeutigam et al., 2008), but was limited to investigations of spectral power and only in the gamma band. Spectral analyses at lower frequencies are warranted, however, as the N400 effect has been associated with increased theta power, as well as with gamma-band effects (Maguire & Abel, 2013).

The current study uses EEG spectral analysis (power and coherence) and ERP analysis to examine patterns of neural activity and connectivity during semantic processing in high-functioning individuals with autism (HFAs) and normal controls (NCs).

Hypotheses:

- ASD will show a reduced or absent N400 compared to NCs
- The N400 will be associated with increased theta power, which may be reduced or absent in HFAs in accordance with N400 differences
- HFAs will show reduced EEG coherence compared to NCs between left fronto-parietal electrode pairs during, or just before, the N400 window

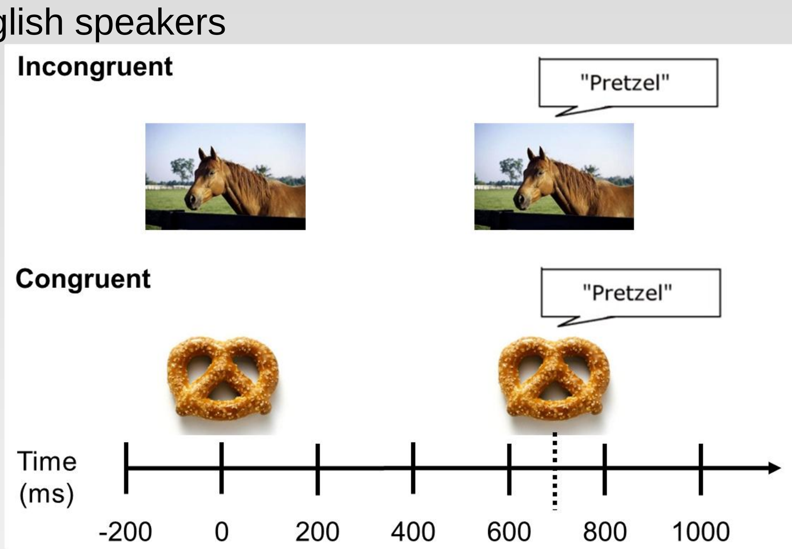
Methods

Participants

- 11 HFAs; mean age 29 years (SD = 14); 10 males, 1 female; 7 Caucasian; 2 African American; 1 Asian; 1 Hispanic
- 11 NCs matched on age and sex; mean age 28 years (SD = 12); 10 males, 1 female; 7 Caucasian; 3 African American
- All right-handed native English speakers

Procedure:

- Picture-word incongruency paradigm: 80 high-frequency spoken words paired with 80 pictures.
- Each picture presented twice, once with a congruent and once with an incongruent spoken word



EEG Data Acquisition and Preprocessing

- EEG recorded at 250 Hz using an Electrical Geodesics Inc. GES 300 EEG System with 256-channel Hydrocel Geodesic Sensor Nets and NetStation version 4.3
- Epochs time-locked to picture stimulus
- Motion and eye movement artifacts corrected using ICA decomposition

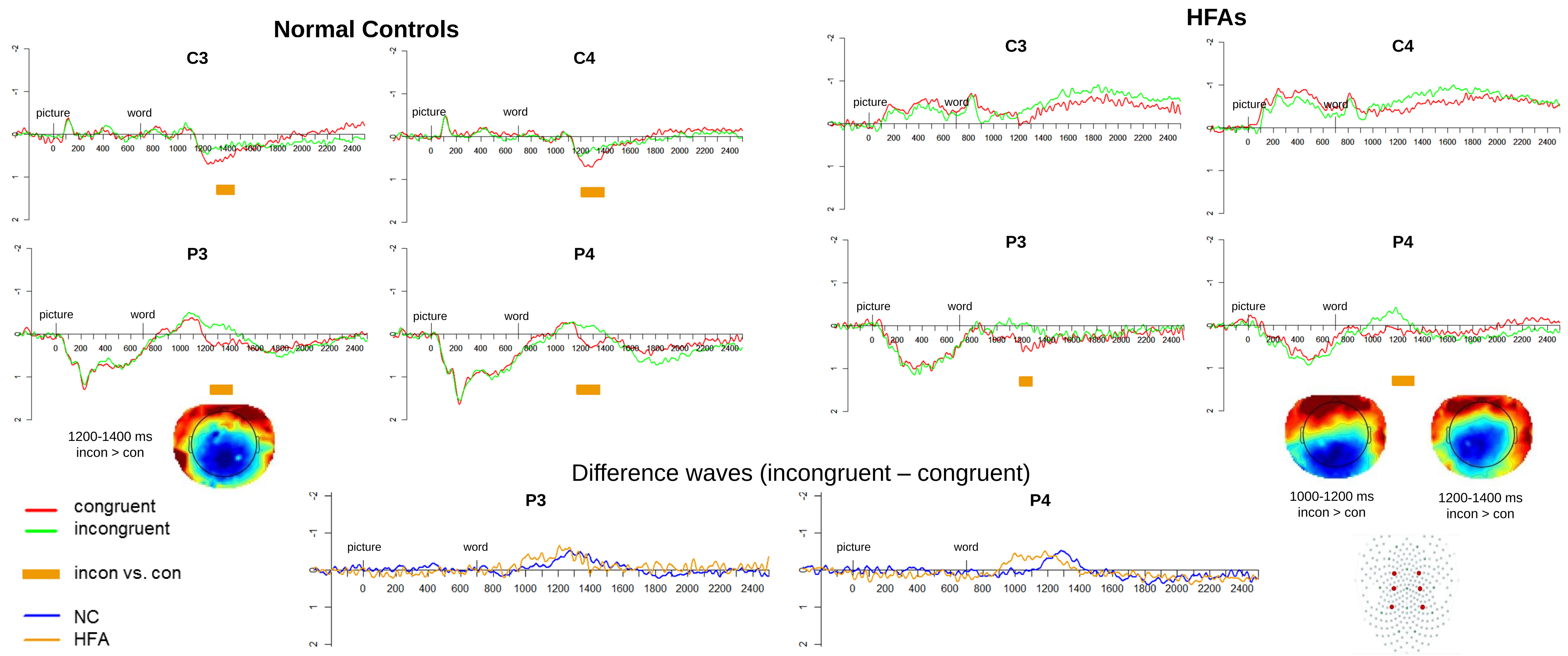
Time-frequency analysis:

- Morlet wavelet of 2 cycles with expanding factor of 0.5 and Hanning taper
- Frequencies 2-50 Hz (delta to gamma)

Results

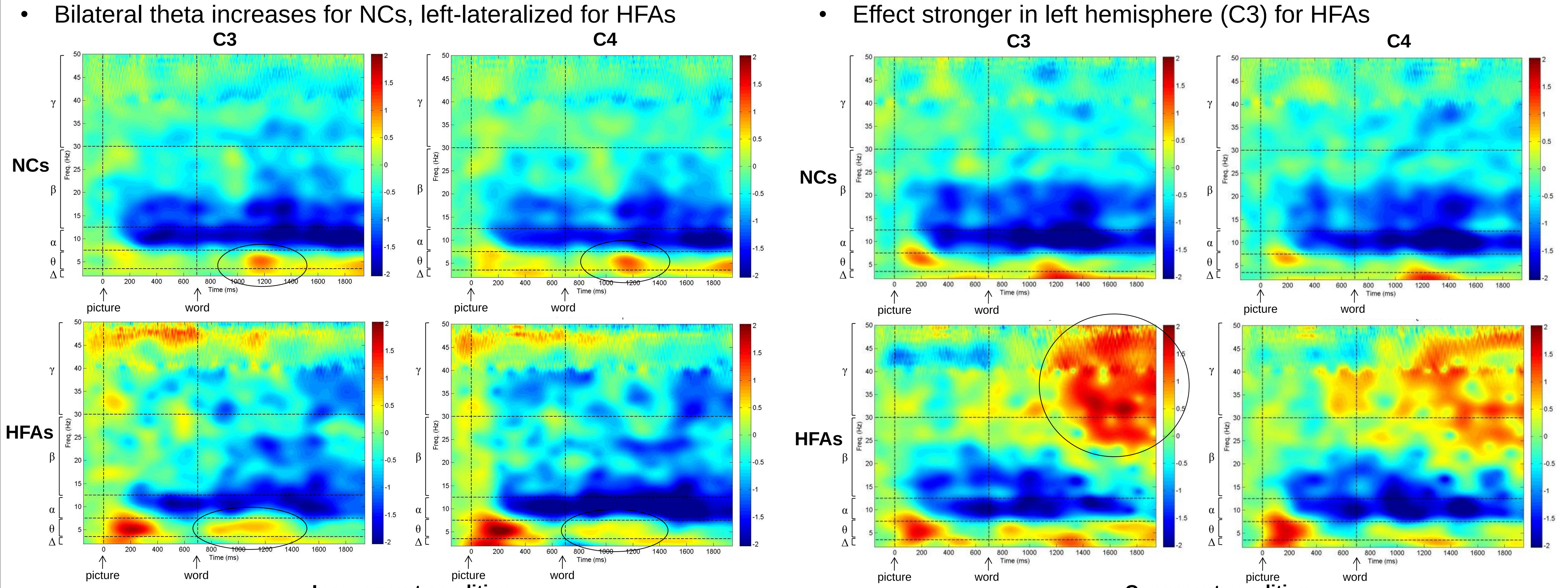
ERPs

- Centro-parietal N400 effect for both groups; slightly earlier onset and more sustained effect for HFAs than for NCs



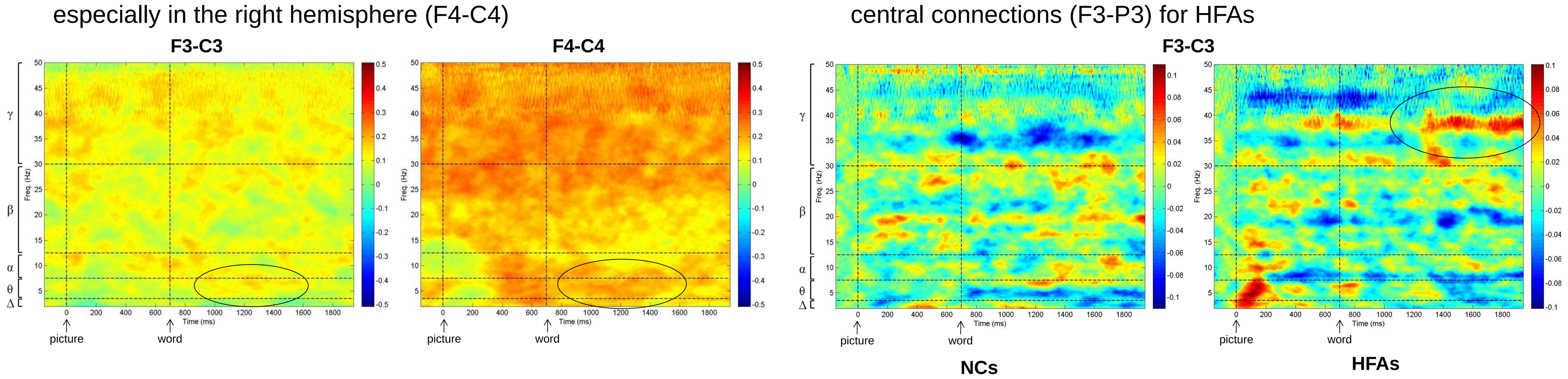
Power

- In the **incongruent condition**, both groups show increased theta power just before N400 onset
- Bilateral theta increases for NCs, left-lateralized for HFAs
- In the **congruent condition**, increased gamma activity starting 700 ms after sound presentation for HFAs but not NCs.
- Effect stronger in left hemisphere (C3) for HFAs



Coherence

- Theta power in incongruent conditions is associated with greater coherence in theta-band for NCs in fronto-central connections, especially in the right hemisphere (F4-C4)
- Gamma power in congruent conditions is associated with greater gamma-band coherence in left-hemisphere fronto-central connections (F3-P3) for HFAs



Discussion

ERPs:

Both NCs and HFAs showed an N400 effect, although it was earlier and more sustained for HFAs. This does not support previous literature, which had found no N400 for HFAs in a picture-spoken word semantic integration task (e.g. McCleery et al., 2010).

Spectral analyses:

In **incongruent conditions**, both groups showed increased power in the theta band starting just before N400 onset. Theta power increases were larger for NCs. NCs also showed a bilateral theta power increase for NCs, whereas this effect was left-lateralized for HFAs.

Theta power changes were associated with reduced fronto-central theta coherence for HFAs vs. NCs for left (F3-C3) and especially right (F4-C4) hemispheres.

These results suggest reduced theta power and reduced fronto-central connectivity for HFAs compared to NCs, especially in the right hemisphere, during the N400 window.

In **congruent conditions**, HFAs showed an increase in gamma-band power starting approximately 600 ms after word presentation; this effect was absent in NCs. This supports previous findings that HFAs show stronger gamma-band increases than NCs (Braeutigam et al., 2008).

Gamma power changes have been associated with the predictability of language, showing larger power increases in response to highly-predictable, congruous semantic contexts (Maguire & Abel, 2013; Wang et al. 2012). The larger gamma activity in HFAs could suggest that they were actively predicting the picture name in preparation for the spoken word. This could explain why HFAs also showed an N400 effect: because they were given explicit instruction to attend to the semantic relationship between picture-word pairs, HFAs may have developed a compensatory strategy that allowed them to perform similarly to NCs (Koolen et al., 2014).

This change in gamma power was associated with increased gamma-band coherence in left fronto-central (F3-C3) connections for HFAs. This suggests that language networks in the left hemisphere may be intact in ASD, but may be recruited in ways that differ from NCs.

The results presented here are preliminary; we are still in the process of collecting data. Currently there are no statistically significant group differences in the spectral analyses due to the strict corrections needed for multiple comparisons and due to the lack of power from so few subjects; however, we expect that this will change with additional data.

Conclusions

Overall, these results suggest differences in event-locked power and coherence during semantic processing in HFAs compared to NCs.

References

Braeutigam, S., Swithenby, S.J., & Bailey, A.J. (2008). Contextual integration the unusual way: a magnetoencephalographic study of responses to semantic violation in individuals with autism spectrum disorders. *European Journal of Neuroscience, 27*, 1026-1036.
Jones, T.B., Bandettini, P.A., Kenworthy, L., Case, L.K., Milleville, S.C., Martin, A., Birn, R.M. (2010). Sources of group differences in functional connectivity: an investigation applied to autism spectrum disorder. *Neuroimage*, 49, 401–414.
Just, M.A., Cherkassky, V.L., Keller, T.A., Minshew, N.J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127, 1811–21.
Koolen, S., Vissers, C.Th.W.M., Egger, J.I.M., & Verhoeven, L. (2014). Monitoring in language perception in high-functioning adults with autism spectrum disorder: Evidence from event-related potentials. *Clinical Neurophysiology, 125*(1), 108-123.
Kutas, M. & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.
McCleery, J.P., Ceponiene, R., Burner, K.M., Townsend, J., Kinnear, M., & Schreibman, L. (2010). Neural correlates of verbal and nonverbal semantic integration in children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *51*(3), 277-286.
Tager-Flusberg H, Paul R, Lord C: Language and communication in autism. In *Handbook of Autism and Pervasive Developmental Disorders*. 3rd Edition. Edited by Volkmar F, Paul R, Klin A, Cohen D. New York: John Wiley & Sons; 2005:335–364
Wang, L., Zhu, Z., & Bastiaansen, M. (2013). Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology, 3*(187), 1-12.

ACKNOWLEDGEMENTS

This work is supported by the Therapeutic Cognitive Neuroscience fund and The Benjamin and Adith Miller Family Endowment on Aging, Alzheimer's Disease, and Autism Research.

Cognitive Neuroscience Society; Boston, MA; April 5-8, 2014
ecoderr1@jhmi.edu

Appendix 6

Coderre, E., Cherenok, M., O'Grady, J., Bosley, L, Gordon, B., & Ledoux, K. (2015, September). Event-Related Potentials as Implicit Measures of Vocabulary in Individuals with Autism. Poster presented at the American Neurological Association's 2015 Annual Meeting, Chicago, IL.

Event-Related Potentials as Implicit Measures of Vocabulary in Individuals with Autism

Emily Coderre[1], Mariya Chernenok[1], Jessica O'Grady[1], Laura Bosley[1], Barry Gordon[1,2], & Kerry Ledoux[1]

[1]Cognitive Neurology/Neuropsychology, Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD

[2]Department of Cognitive Science, The Johns Hopkins University, Baltimore, MD

Introduction

Assessments of the cognitive operations responsible for language are typically quantified by measuring overt behaviors such as response time or verbal reports. However, such explicit measures assume an understanding of task goals and an ability to execute the required response. In certain populations, such as non- or minimally-verbal low-functioning individuals with autism (LFAs) in whom such measures might be difficult or impossible to obtain, implicit measures of cognitive abilities that do not require explicit understanding and cooperation are essential.

Event-related potentials (ERPs) can serve as implicit measures of vocabulary knowledge. The amplitude of the N400 ERP component is influenced by the ease of semantic integration and is reduced to stimuli that are semantically congruent (such as matching pairs of pictures and words), which are easier to integrate relative to those that are incongruent (such as mismatching pairs, which are more difficult to integrate; Kutas & Federmeier, 2011). This modulation by congruency, or "N400 effect", is limited to the individual's vocabulary range: no such effect occurs for unknown words, for which prior knowledge cannot help ease integration (Connolly & D'Arcy, 2000).

In recent work, we have shown that ERPs can be used to estimate vocabulary knowledge in normal adults (Ledoux et al., 2015). In a picture-word congruity paradigm, an N400 effect was observed for high-frequency 'known' words but not for low-frequency 'unknown' words, suggesting that the N400 effect can reliably estimate vocabulary knowledge in a population of normal adults.

Although ERPs hold potential for cognitive assessment in the absence of behavioral responses, the utility of these measures in individuals with autism has not been determined. Here we investigate whether ERPs can serve as within-subject measures of vocabulary knowledge in in individuals with autism with a range of functioning levels.

Methods

Participants

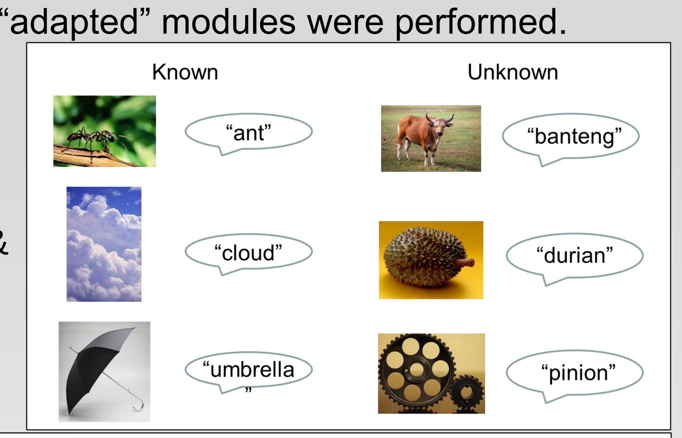
- 24 participants with autism; mean age 29 years (range 15-66); 23 males; 19 Caucasian, 1 African American, 3 Asian, 1 Hispanic.
- 9 participants were enrolled in adult or educational programs specific to assisting individuals with autism and required direct 24-hour support staff and/or parental supervision.

Neuropsychological Testing

- Receptive language abilities: Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007)
- Verbal and non-verbal intelligence: Kaufman Brief Intelligence Test, Second Edition (K-BIT-2; Kaufman & Kaufman, 2004)
- Autism symptoms: Autism Diagnostic Observation Schedule (First Edition (ADOS-1) or Second Edition (ADOS-2), depending on the version current at the time of testing; Lord et al. 2000).
  - For 3 participants there was no appropriate module of the ADOS, as currently no modules address nonverbal adolescents or adults. For these participants, "adapted" modules were performed.
- Some participants were unable to complete behavioral testing due to lack of compliance or inability to understand task instructions

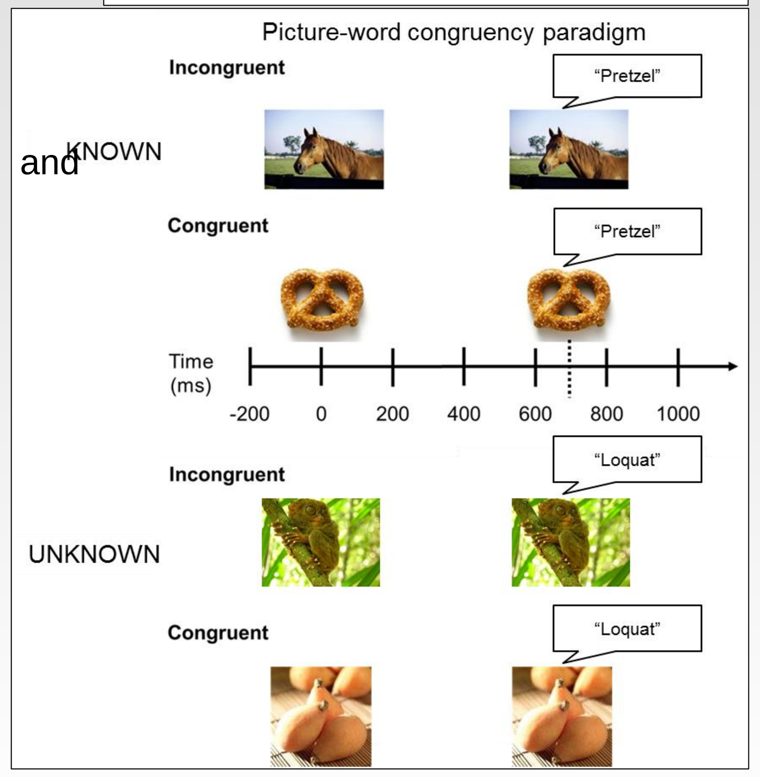
Stimuli

- 80 high-frequency words (average SubtlexUS (Brysbaert & New, 2009) log10 frequency rating = 3.14, *SD* = 0.6). Because of their high frequency, these words were expected to be 'known' to participants
- 80 low-frequency words (average SubtlexUS log10 frequency rating = 0.85, *SD* = 0.5). Because of their low frequency, these words were expected to be 'unknown' to participants
- Corresponding high-resolution color photographs auditory recordings



Procedure

- Picture-word congruency paradigm: each picture presented twice, once with congruent and once with incongruent word pairing

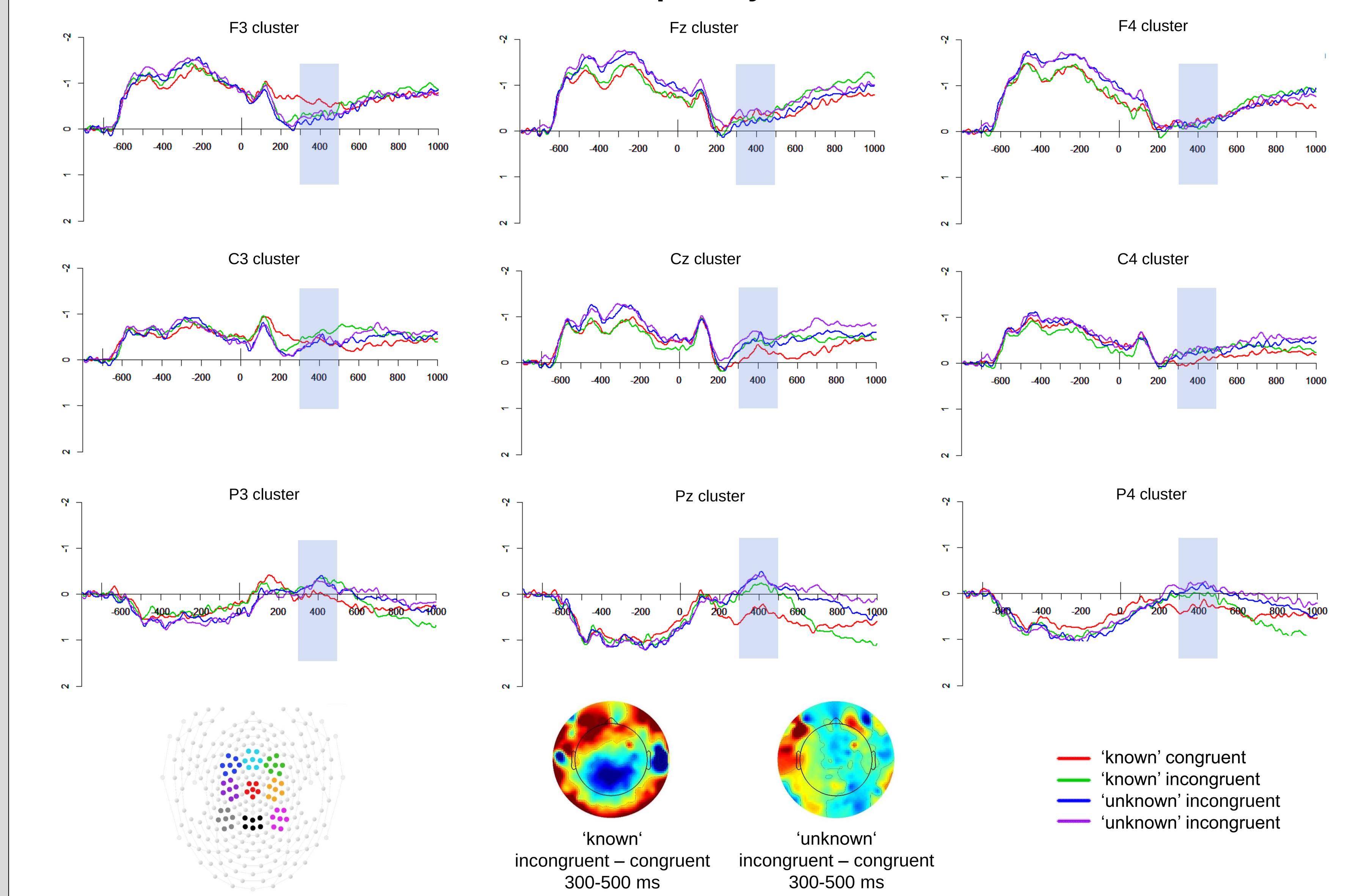


EEG Data Acquisition and Preprocessing

- EEG recorded at 250 Hz using an Electrical Geodesics Inc. GES 300 EEG System with 256-channel Hydrocel Geodesic Sensor Nets and NetStation version 4.3
- Bandpass filter 0.1-30Hz. Motion and eye movement artifacts corrected using ICA decomposition
- Electrodes grouped into 9 clusters for analyses

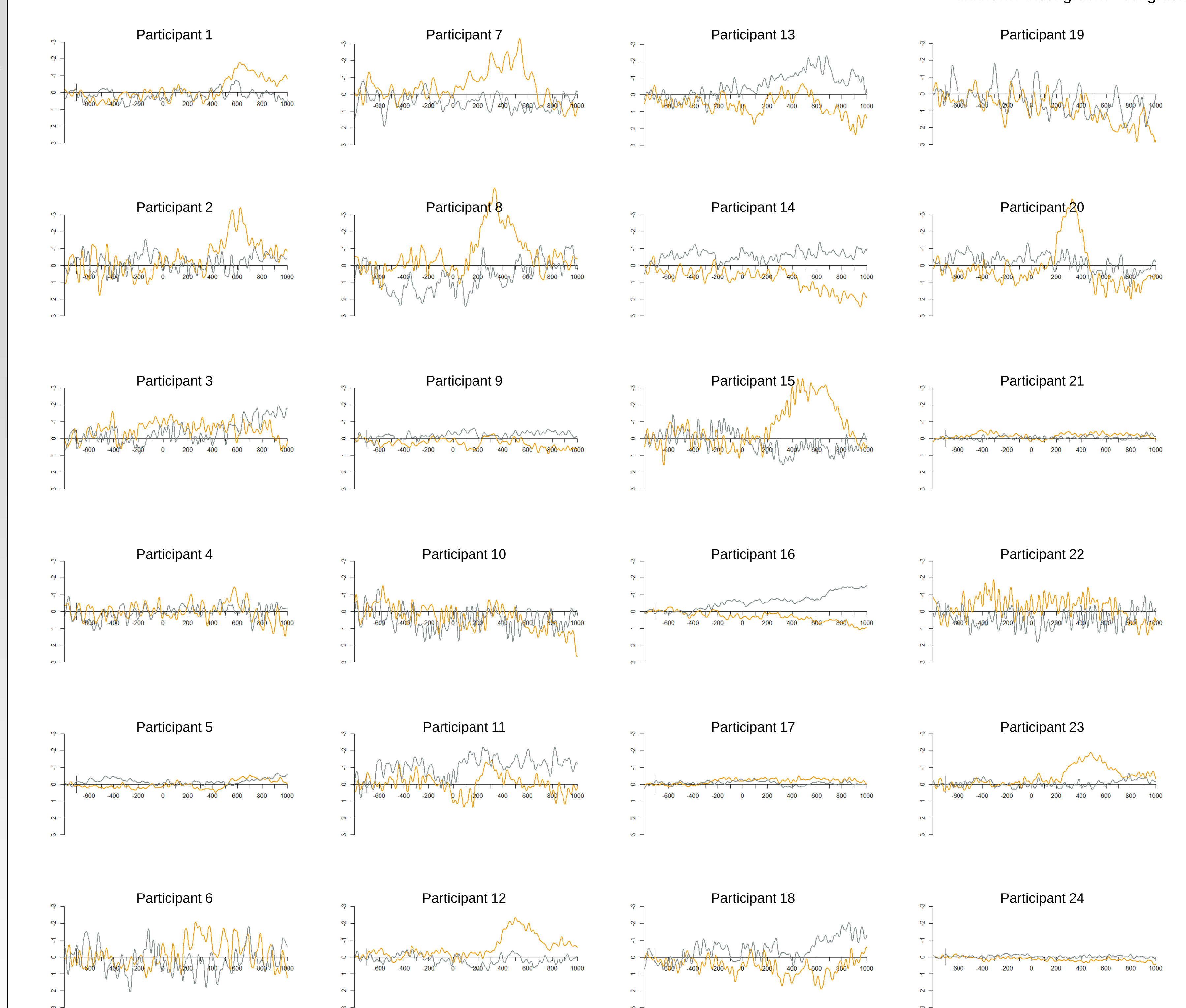
Results

Group analysis

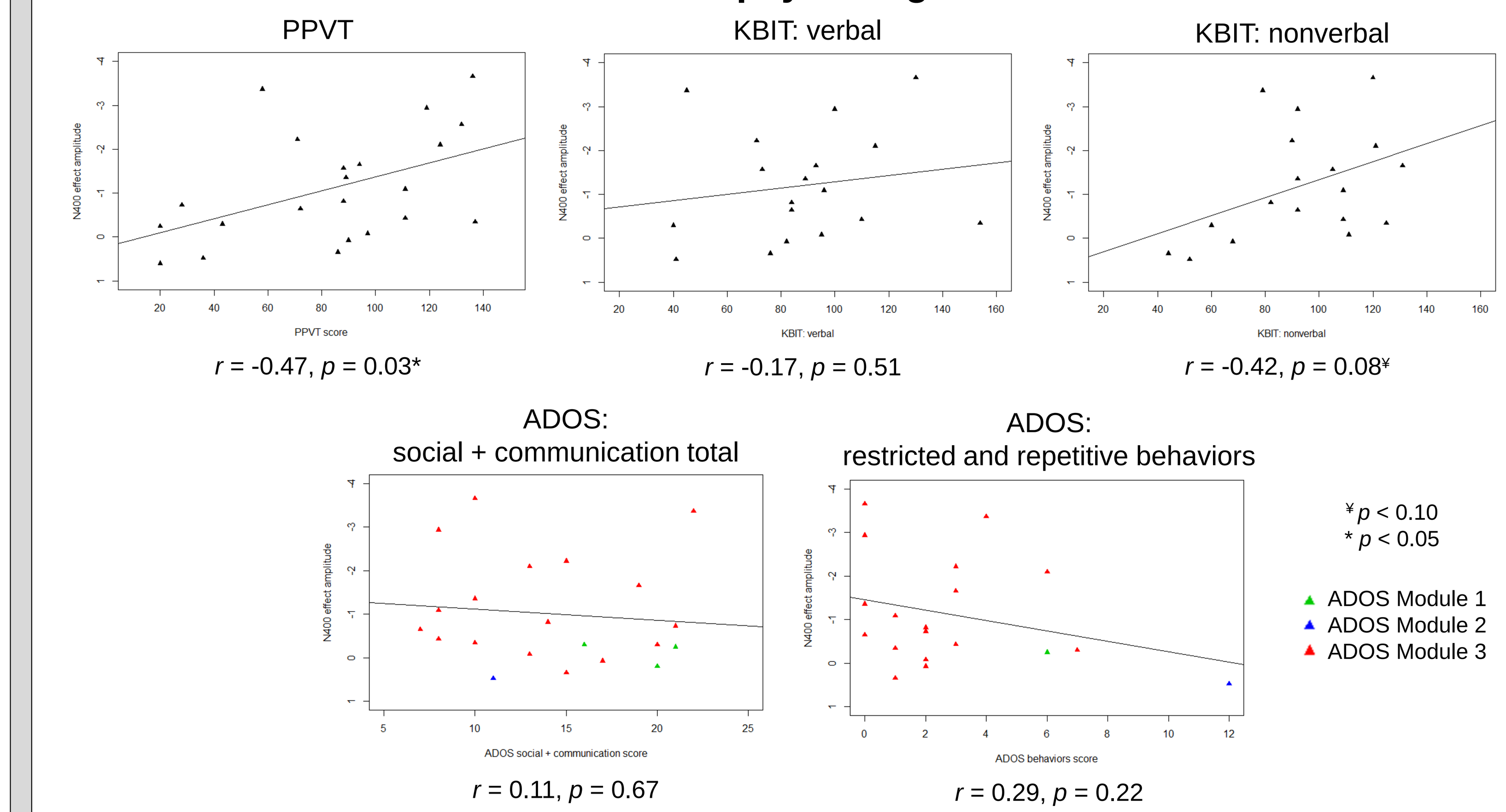


- A 2 (knowledge: known/unknown) x 2 (congruency: congruent/incongruent) x 3 (site: frontal/central/parietal) x 3 (laterality: left/midline/right) repeated-measures ANOVA was run on the average amplitude from 300-500 ms (blue shaded areas).
- A significant four-way interaction ($F(4,92) = 2.58$, $p < 0.05$) arose from trends towards significant interactions of knowledge and congruency at left parietal (P3 cluster; $F(1,23) = 3.35$, $p = 0.08$) and midline parietal (Pz cluster; $F(1,23) = 3.56$, $p = 0.07$) sites.
- At both P3 and Pz clusters an N400 effect occurred in known conditions ($p < 0.05$) but not unknown conditions (all p's > 0.64).

Individual effects at Pz cluster



Correlations with neuropsychological assessments



- Each subject's average N400 amplitude was calculated by finding the peak negative amplitude of the known condition difference wave at the P3 cluster, then averaging over a window 50 ms before and after the peak.
- Pairwise Pearson correlations between N400 effect magnitude and behavioral scores showed significant correlations between the N400 magnitude and PPVT scores, with a trend between N400 magnitude and nonverbal KBIT scores.

Discussion

In the group analysis, "known" words elicited an N400 effect over centro-parietal scalp, whereas there was no such effect for "unknown" words. These findings replicate the results observed in normal adults by Ledoux et al. (2015) and demonstrate that ERPs can serve as within-subject measures of vocabulary knowledge in in individuals with autism across a range of functioning levels.

Correlational analyses showed a significant correlation between PPVT scores and N400 effects, such that participants with better vocabulary abilities (larger PPVT scores) showed larger N400 responses. This correlation replicates previous findings in the literature (D'Arcy et al., 2003) and suggests that the N400 response is accurately capturing vocabulary knowledge without reliance on behavioral measures.

The individual data demonstrate significant heterogeneity among the participants. While some had large N400 responses in "known" words, others showed little difference between congruent and incongruent stimuli in either "known" or "unknown" words. This variability suggests that the N400 may be better suited as an implicit estimate of vocabulary knowledge in individuals with autism who show larger effects. Factors such as the ability to tolerate the EEG net and the number of sessions required to obtain enough clean data should also be considered.

Conclusions

Overall, the N400 distinguished between "known" and "unknown" vocabulary in individuals with autism and correlated with receptive language abilities, although there was significant individual variation. Despite the heterogeneity inherent in autism, ERPs can serve as implicit measures of vocabulary in this population, and hold especially strong potential for language assessment in low-functioning individuals.

References

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Connolly, J.F., & D'Arcy, R.C.N. (2000). Innovations in neuropsychological assessment using event-related brain potentials. *International Journal of Psychophysiology, 37,* 31-47.

D'Arcy, R. C. N., Marchand, Y., Eskes, G. a, Harrison, E. R., Phillips, S. J., Major, A., & Connolly, J. F. (2003). Electrophysiological assessment of language function following stroke. *Clinical Neurophysiology*, *114*(4), 662–672.

Dunn, L., & Dunn, D. (2007). Peabody Picture Vocabulary Tests (4th edition). *American Guidance Service*. Circle Pines, MN: American Guidance Service.

Kaufman, A., & Kaufman, N. (2004). Kaufman Brief Intelligence Test (2nd edition). Circle Pines, MN: American Guidance Service.

Kuipers, J.R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, 5, 61.

Kutas, M. & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Ledoux, K., Coderre, E.L., Bosley, L., Buz, E., Gangopadhyay, I., & Gordon, B. (2015). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavior Research Methods.* doi 10.3758/s13428-015-0571-6.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Acknowledgements

We would like to thank the Linwood Center for their involvement as a research site and all of the participants and their families and/or caregivers. This research was supported by the Nancy Lurie Marks Foundation; the Department of Defense Autism Research Program AR093137; The Therapeutic Cognitive Neuroscience Fund; and the Benjamin and Adith Miller Family Endowment on Aging, Alzheimer's, and Autism Research.

Presented at the 2015 annual meeting of the American Neurological Association, September 27-29, Chicago, Illinois